Prediction in Gaze and Saccade Control

Christopher M. Brown

Technical Report 295
May 1989

UNIVERSITY OF
ROCHESTER
COMPUTER SCIENCE

Prediction in Gaze and Saccade Control

Christopher Brown
Department of Computer Science
University of Rochester
Rochester, NY 14627

May 13, 1989

Abstract

Animate vision systems, biological or robotic, employ gaze control systems to acquire, fixate, and stabilize images. Gaze control and other sensorimotor skills are complicated by two main problems: time delay and the interaction of control subsystems. One solution is open-loop control. The other, explored here, is control with prediction. The goal is to build robust gaze control behavior from cooperating lower-level "visual reflexes". Solutions are explored through simulation incorporating ten primitive control cabablities, more or less comparable to subsystems in primate gaze and head control. Versions of several of the subsystems have been implemented on a binocular robot. Signal synthesis adaptive control with Smith prediction is the basic paradigm, implemented with kinematic simulation of the agent and optimal filtering to predict world state. The task of rapid gaze shift illustrates several issues. Predictive methods appear useful for artificial gaze control systems and often produce results consistent with physiological data.

Keywords: dynamic Systems [illegible] (KR)

Acknowledgements

The work was supported by the DARPA U.S. Army Engineering Topographic Laboraties Grant DACA76-85-C-0001, the Air Force Systems Command (RADC, Griffiss AFB, NY) and Air Force OSR Contract F30602-85-C-0008, which supports the Northeast Artificial Intelligence Consortium. The University of Rochester College of Arts and Sciences provided sabbatical leave support. The assistance of Mike Brady, Donald Weir, Joe Mundy, Hugh Durrant-Whyte, Andrew Blake, Ron Daniel, and the staff of the Robotics Research Group of the University of Oxford is gratefully acknowledged.

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER 295	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Prediction in Gaze and Saccade Control		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Christopher M. Brown		8. CONTRACT OR GRANT NUMBER(s) DACA76-85-C-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department 734 Computer Studies Bldg University of Rochester, Rochester, NY 14627		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS D. Adv. Res. Proj. Agency 1400 Wilson Blvd Arlington, VA 22209		12. REPORT DATE May 1989
		13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) US Army, ETL Fort Belvoir, VA 22060		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution of this document is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Predictive control, active vision,
gaze control, Dynamic Systems estimation

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Animate vision systems, biological or robotic, employ gaze control systems to acquire, fixate, and stabilize images. Gaze control and other sensorimotor skills are complicated by two main problems: time delay and the interaction of control subsystems. One solution is open-loop control. The other, explored here, is control with prediction. The goal is to build robust gaze control behavior from cooperating lower-level "visual reflexes". Solutions are explored through simulation incorporating ten primitive control

20. ABSTRACT (Continued)

capabilities, more or less comparable to subsystems in primate gaze and head control. Versions of several of the subsystems have been implemented on a binocular robot. Signal synthesis adaptive control with Smith prediction is the basic paradigm, implemented with kinematic simulation of the agent and optimal filtering to predict world state. The task of rapid gaze shift illustrates several issues. Predictive methods appear useful for artificial gaze control systems and often produce results consistent with physiological data.

1 Gaze Control in Animate Vision

1.1 Gaze Control

One of the goals of artificial *animate vision* [4,5] is to design a systems architecture in which multiple objectives (such as moving and observing) can proceed in parallel. One common premise is that cognitive processes at high abstraction levels can rely on a hierarchy of lower-level "skills", "reflexes", and active vision capabilities that autonomously keep the robot out of trouble and perform generally useful vision computations [10]. Another premise is that if an agent interacts with the world and has active control over and perception of its own state (*proprioception*), many problems in perception, planning, and acting become easier. A *gaze control* system manages several basic, interacting head, eye, and even body motion capabilities, with the aim of supporting purposeful (or default) activity. For example, the ability to acquire an object visually is basic. *Acquisition* can be reflexive, in response to a stimulus deemed interesting, or under control of a higher level engaged in planning or acting. Another basic visual ability is to *pursue* or track an object moving relative to the observer: Stabilization of the image on the sensor is necessary for high resolution imaging, and the resulting proprioception (i.e. motor commands that effected the tracking) provides information about object motion. Another example is automatic vergence of a binocular eye system, which provides a clue to object range and is helpful for obtaining stereo correspondence. Many gaze-control issues become even sharper if the imaging system has both foveal and peripheral vision, as do primates or robots equipped with wide-angle and telephoto cameras. A gaze control system has two main *technical problems:* the *interaction of component subcontrols* and *delay.* These problems are common to much fast sensorimotor skilled behavior in biological and current robotic systems (locomotion, manipulation, navigation); techniques for solving them can find wide application.

Gaze control mechanisms have long been studied in biological systems (there are extensive references in [21,8,33]). Much of the work concentrates on how biological systems solve the two basic technical problems mentioned above. This paper investigates predictive mechanisms as a solution for the problems primarily in a robotic context, but occasionally relates the results to some findings and theories from primate gaze control. This work is in the spirit of the *science of the artificial*, not the *science of the natural* – physiologists and engineers, for different reasons, may find such comparisons unhelpful. The hope, however, is that *data from successful (biological) systems performing gaze-control tasks can provide the (artificial) animate vision system designer with ideas and benchmarks, despite clear and significant differences in underlying implementation.*

As robotic technology approaches biological performance [2,4] the potential for sharing descriptive and computational models between the two fields increases. Of course, some robotic physical capabilities and tasks allow or require sensorimotor capabilities that have little in common with biological systems. The gaze control task mainly considered here is general and simple: rapid gaze shift using head and eye motions.

1.2 Delays and Interactions

It is well known that delays in exerting control or acquiring data can severely degrade or destabilize the performance of feedback control systems, as can unmodelled effects that change the system, such as another control system acting on the same *plant.* Gaze control systems have several controls operating at once, and biological and robotic systems have delays for the same reasons: computation and transmission. Robotic delays depend critically on the software and hardware, and can have very unpleasant properties (such as unpredictable magnitudes) without special engineering attention. Biological delays are more predictable but are long with respect to the behavior of the system. Human saccadic and pursuit systems have computation delays of about 120ms and 50 ms respectively, and both have peripheral (transmission) delays of about 80 ms. Saccades last only 50–100 ms, and the pursuit system brings the eyes up to speed in about 120ms. It is easy to demonstrate that these delays are incompatible with pure negative-feedback control schemes.

There are two main solutions: *avoid feedback* or *model delays and disturbances.* The former method is advocated in [35], and has been implemented in a robotic controller [15]. The idea is to use positive feedback to cancel negative feedback, thus producing an open loop system. The main disadvantage is the loss of robustness under disturbances and plant variations. Parameter-adaptive control can be added to adjust system parameters over a longer timescale to reduce systematic errors.

The other approach, more common in engineering applications, is to use *predictive* and modeling techniques to anticipate the state of the plant, its input, and indeed the world [24]. *Smith's principle* [37,38] is the basic tenet that the desired output from a controlled system with delay T is the same as that desired from the delay-free system, only delayed by T. The principle leads to several techniques for controlling delayed systems (see Section 4). Smith's principle may be coupled with *signal synthesis adaptive control* [3], which predicts object motion to allow more accurate responses. Kinematic and dynamic models for plant prediction can be known apriori or derived from learning and used to replace feedback [23].

The solution described here uses Smith prediction to integrate multiple controls with delays. It uses signal synthesis adaptive control with flexible and general techniques of *kinematic simulation* to predict the state of the plant and variance-minimizing *optimal filtering* to predict the state of the world. At least in simulation, the resulting predictions have three effects. *Delays are overcome, interactions are overcome, and performance is improved.* Predictive techniques seem to form a sound basis for the design of integrated, high performance sensorimotor systems.

2 The Animate System

The control algorithms are based on a general model that has been implemented in simulation, and both model and simulation are meant to capture the relevant aspects of a

* * * Figure 1 about here * * *

Figure 1: *The robot head and its supporting arm.*

* * * Figure 2 about here * * *

Figure 2: *The laboratory computer organization. The 24 node Butterfly will replace the host Sun, and faster interfaces to the Puma and MaxVideo are being installed.*

robotic system which, to fix ideas, is described first.

2.1 The Rochester Robot

Research in artificial animate vision has been made possible by recent technical advances in real-time computer vision and control. The area still includes a wide spectrum of problems.from hardware design through software for parallel systems support and applications, to the integration of heterogeneous computers, sensors, and effectors into behaving systems. Many laboratories are developing similar systems to investigate these issues and to take advantage of the new technologies, and it seems that complex visuo-motor systems will be the rule in the robotics laboratory of the near future. The usual current setup has controllable sensors (often bi- or trinocular TV cameras, perhaps sonar) and powerful parallel computation, including frame-rate image analysis hardware. Sometimes the sensors are mounted on a roving cart.

Rochester's binocular robot head is mounted on a six degree of freedom arm (Fig. 1). The two cameras are on a common tilt platform, and have independent pan axes. The hardware is capable of motions comparable to primate performance (about 1 m/sec head velocity with less than 1 mm. positioning accuracy, and 300 degrees/sec camera rotations with .14 degree positioning accuracy). The camera controllers are capable of supporting full-speed gaze-shifts to random directions at a rate of 5/s. The camera and robot controllers support several types of motion, but the frequency responses of system components under various forms of control have not been measured. The aperture, focal length, and focus of the cameras are not yet controllable. The video output is processed by a Datacube MaxVideo pipeline-parallel image processing system that can do many low and intermediate-level vision operations at 30Hz (video frame rate). The host computer for the system is currently a Sun/3 computer but will soon be a 24-node Butterfly Parallel Processor (BPP). Each node of the BPP is an M68020 processor with M68881 floating point coprocessor and 4MB of memory. There is fast switch allowing the processors to share memory. The plan is to implement the control algorithms in the BPP, using multiple nodes as necessary for speed, with the Datacube furnishing real-time input. A real-time package for the Psyche operating system will support the applications. LISP planning programs can communicate with robot applications code over the ethernet. Fig. 2 shows the current hardware organization.

In the absence of proprioceptive devices like shaft encoders to measure robot state directly, the current state of the physical robot and cameras is taken to be the current values reported by their controllers. For various reasons this method is less reliable for roving carts. The frame-rate image processor can produce fairly sophisticated current states of the image such as depth maps or the centroids of multiple blobs. The major uncertainty in predicting future states arises not from mechanical imprecision but from the unpredictability in timing of the control commands caused by software delays.

Recent work with the Rochester Robot (RR) produced several implementations of potential basic components of a real-time gaze-control system [4,12,6,32]. These components included basic capabilities of *object tracking, rapid gaze shifts, counteracting head motions with compensating eye motions, verging the cameras, binocular stereo*, and *kinetic depth.* For instance, pursuit is accomplished with a special-purpose board that correlates the image with an 8 × 8 template of the pattern to be tracked, and a board that converts over-threshold correlation peaks to image (x,y) coordinates, which are then converted to camera pan and tilt commands. Vergence signals arise from a cepstral filter (similar to phase correlation) implemented with the integer fast fourier transform on a digital signal processing chip, producing a disparity measure that is then converted into a camera pan command. These several capabilities should cooperate to accomplish tasks. Some couplings between capabilities are obvious: for binocular vision, vergence can be closely (perhaps even mechanically) coupled with focus, and indeed it is known that there is intimate bidirectional coupling between primate vergence and accommodation [28]. What is needed is a way to integrate the several capabilities (and others) smoothly for a range of tasks useful for perception, navigation, manipulation, and in general "survival".

2.2 The Simulator

The gaze control design described here uses a simulator to demonstrate results. and more fundamentally as a basic part of the predictive control system. The simulator should apply to a large class of kinematic, imaging, and control systems including the RR. It incorporates head geometry and models angular velocity controls acting with delays corresponding to delays caused by transmission and computation. In the work reported here delay arises from the controllers, not the sensors, as is the case with many of the simpler algorithms implemented on modern vision hardware. In general there is sensor delay as well, and dealing with it requires straightforward extensions of the techniques discussed below. In [11] there were five controls, in the current work there are ten, (depending on how variants are counted). Some controls are reminiscent of primate capabilities, some are not, but they span a range of capabilities of varying sophistication and seem to be reasonable building blocks for more complex gaze control skills.

The simulated robot has a HEAD (a coordinate system) that can be translated and rotated under velocity control. The modeled cameras are at the ends of kinematic chains attached to the HEAD origin, with their individual pan and common tilt velocities controllable. The *object* of visual interest is modeled as a single point in three dimensions, and its image as a retinal point. This idealized percept ignores many of the phenomena that cause problems for low-level computer vision but also provide information to both artificial and biological systems (e.g. motion blur, optic flow (retinal slip). and moving edges). The robot state also includes the position of the object's image (under perspective projection) in both cameras, its incremental motion since the last simulated instant (its instantaneous optic flow), and three-dimensional object position information (say from kinetic depth, binocular stereo, a sonar sensor, or even memory).

The geometric configuration, speeds, delays, and time constants of the simulation can be altered to conform with a variety of robotic or biological specifications. The simulation of eye movements by pan and tilt (not allowing rotations about the optic axis) seems reasonable for primate modeling. The kinematic simulation is satisfactory at a certain level of abstraction. The RR's existing commercial controllers hide the dynamics from applications software, and dynamics do not seem necessary to model most aspects of primate eye movements (though may be necessary for a responsible attack on head movements [33].) The abstraction of the six-link robot arm to the model HEAD coordinate system is in fact supported by software in the robot controller, which provides commands that allow motions of the end of the arm (bearing the head) in the HEAD coordinate system. A single origin of head coordinates does not do justice to the primate head and neck joints, but is a reasonable approximation and would make dynamics easy to compute. The discrete-time nature of the simulation is not traditional but allows unlimited freedom in control law design.

For this paper the simulator was geometrically configured like the RR, and other parameters were set to illustrate the phenomena of interest. The main task studied in this paper is quick gaze shifts to static or slowly moving objects, so the graphs showing system performance in this paper (and several of those in [11]) represent the step response of the simulated system. Further details on the simulation implementation and some of the control systems appear in [11] and Section 3.1.

3 The Gaze Controls

3.1 Basic Gaze Controls

This section (and the rest of the paper) deals, unless otherwise explicitly noted, with simulated capabilities and occasional psychophysical and physiological data. It gives a brief summary of [11] and relates the goals of this paper to the previous work. Many of the *technical assumptions* in the previous and present work are really *hypotheses*: they are part of the process of discovering how to build a working system. In [11] there were five control systems. Four of them resembled capabilities that had already been implemented on the RR, and are functionally similar enough to primate capabilities that we borrowed the biological names.

1. **Saccadic:** a *closed-loop* eye control to produce rapid gaze shifts, "sampled" at a relatively coarse interval with respect to the discrete simulation "tick", and driven from positional retinal error (distance from retinal origin). At each activation it

produced a sequence of maximum-speed pan and tilt velocity commands that was calculated from approximate kinematics to center the image in the retina.

2. **Smooth pursuit:** a proportional, integral, derivative (PID) eye control driven by the retinal *positional* error of the object's image, to follow objects in motion relative to the eyes. This and the following controls are "continuous" in that they act at every simulation tick – we use the quoted "continuous" and "sampled" in the sequel to make this distinction in a discrete (sampled-data) implementation.

3. **Vergence:** a PID eye control to reduce disparity between left and right images.

4. **Vestibulo-ocular reflex (VOR):** an open-loop proportional-gain eye control to oppose head motion with contrary eye rotation. Head rotation is easily cancelled, and this capability also includes a version of "otolith-ocular reflex" that compensates for translational head motions using information about object range [9,14]. Its input is not sensory, but an "efferent copy" of the head translation and rotation command: this avoids its interference with smooth pursuit when head compensation is active [40].

5. **Head compensation:** a proportional-gain head-control system driven by head-relative eye position that rotates the head in the direction of eye rotation to keep the eyes centered and away from their mechanical stops. This capability may not mirror any named biological one, but it does reflect a widespread primate tendency to move the head, if allowed, during pursuit and eye saccades.

The robot designer need not copy nature: Pursuit may be implemented with either position or velocity error feedback, maximum angular speeds in saccades and pursuit may be the same, the VOR need not use sensors, feedback can use arbitrary control laws, and so on. The technical choices should be guided by the desired performance of the system being designed (as they presumably were in nature). If (as may well be the case), commonalities between the robotic and natural problems a' one useful level of abstraction outweigh the technical differences at a lower level, then results may perhaps be shared between the fields. In this work the emphasis is on the general problem of cooperation between capabilities: the neurosciences are always improving the characterization of natural systems, and the design details of a robotic system vary with individual systems.

A smooth pursuit control using *velocity error* was written and tested. By most accounts this is more like the primate smooth pursuit system. Since this version of pursuit is insensitive to position error, only the eye saccadic system was available to center objects on the retina. Depending on the low-level vision algorithms, both velocity and position errors could be available (correlation techniques give displacements, blob-finders give position) so a velocity-sensitive pursuit reflex should be an option. A wide-field version of it roughly corresponds to an aspect of the functionality of the "opto-kinetic reflex" in primates (but see [11]).

Left eye dominance was assumed (hypothesized): pursuit and saccade commands were only computed for the left eye, and vergence commands for the right. What tc do in an actual system is a complicated matter. Specific hardware is devoted to implementation of pursuit and vergence, and needless duplication should be avoided. It may be cheaper in hardware to do pursuit with both eyes with no vergence control. Vergence, however, makes eyes less likely to wander off and pursue separate targets.

The control signals for the translational and rotational velocities of the head (six values) and of the pan, tilt velocities (three values) form a control vector. To simulate delays and commands (such as saccades) that affect velocities over more than one simulation instant, the commands inhabit a *pipeline* indexed by time. Commands with delays will take place in the future, and so are inserted (or summed, if more than one control affects the same output) into the pipeline at the time in the future when they will take effect. The vector for the current time is retrieved and its velocities applied to the current state.

The system operated in two decoupled modes, saccadic and pursuit. Saccadic mode overrides pursuit mode: in saccadic mode the eye saccadic and vergence systems operated, in pursuit mode the pursuit, vergence, head compensation, and VOR operated.

As expected, delays and interactions catastrophically degraded performance of a naive implementation in which each controller acted independently and outputs were simply summed at the effectors. Controls with delays cannot be independently-acting "modules". A control needs information about the effects of its own actions to cope with delay, and information about other controls is needed if all delays are not identical. The simulator used to demonstrate the effects of the gaze controls was extended to maintain the state of the animate system through time in another pipeline. Control and state pipelines extend into the future as far as the maximum control delay plus the maximum duration of any command sequence. The predicted state at the proper delay for each control is used by the Smith controller (see Section 4).

Along with a more detailed description of the simulation and controls, and the derivation of the Smith predictor control equations, the earlier paper describes experiments establishing the effectiveness of the control scheme with multiple interacting controls, under various conditions of sensor noise and stochastic control delay disturbances and of systematically inaccurate models of head-eye kinematics and object motion. It also has comparisons with primate systems and with open loop methods of coping with delay.

3.2 Advanced Gaze Controls

The previous work left several issues open, forming the motivation for the work described here.

1. All controls operated from retinal coordinates. Predictions of object position in head or laboratory coordinates are needed to predict retinal images. Head rotations, pans, and tilts all induce camera origin translations due to the geometry of the head, and non-retinal representations are more robust (as when the object temporarily is lost). *Versions of pursuit, vergence, and eye saccadic capabilities are added that use spatial information.*

2. There was no capability for estimating the state of objects moving in LAB (relative motion was produced with a static object and observer motion.) *A pipeline of object state descriptions is maintained using variance-minimizing filters to predict object state from observations.*

3. The head compensation reflex was the only head control. *The system has "head saccade" and "head pursuit" controls.*

4. The eye saccade control algorithm was unsophisticated, and there was no significant head and eye cooperation for quick gaze shifts. *Four fast gaze-shift algorithms involving both head and eyes are investigated.*

5. The two modes of operation (pursuit and saccadic) were simplistically assumed to correspond to inflexible combinations of lower level capabilities. *All controls can be activated and deactivated independently.*

Unlike the previous work [11], this paper is noncommittal on whether the eye saccadic system should control one or both eyes. Binocular eye saccades are easy to implement, but there are many other ways that the eyes can cooperate during gaze shifts, depending on technical considerations such as the possibility of visual computation during the shift, how the vergence control is specified, etc. We likewise shall not address issues of orchestrating the smooth transition between saccadic and pursuit (or other) tasks. Considerable work is still needed on the topic of smooth blending of subcontrols, which forms the foundation of motor skills.

This work shows a mix of *predictive techniques* controlling a system with interacting, delayed controls. Technically, the work required the *extension of prediction to include the world state* and the *extension of the predictions to non-retinal coordinate systems. Optimal filters*, such as the Kalman filter, are a natural and powerful technique to implement prediction of the world state. Prediction is demonstrated to yield *better pursuit performance. Time-optimal eye-head saccades* can be implemented with predictive techniques, while methods based on feedback do not perform as well.

4 Control with Delays

The Smith predictor incorporating the object state predictions is shown in Fig. 3. Its derivation appears, for instance, in [11,24]. The basic idea is to have a zero-delay feedback (path C) based on simulation. The model-delayed simulation data (path B) is compared with the actually-delayed data from the plant (path A): the difference (at D) is zero for perfect simulation, so it provides information about the simulation adequacy and (if slowly varying compared to the control delays) compensates for inaccurate modeling. It presumably can be used to adapt the control or change the model, but the current controller does not so use it.

The work described here uses the following *interacting controls* algorithm [11], which assumes each controller knows its own delay T, and the delays of all the other controllers

* * * Figure 3 about here * * *

Figure 3: *The Smith predictor control. The CONTROL block represents all control systems, and DELAY their independent delays. K1 and K2 are gains to weight the delayed and non-delayed modeled error signals. The $\alpha - \beta$ filters estimate object and image states, and adaptively synthesize the signals for the control system.*

in the set {S} that share an output with it. *Look ahead the maximum delay M of any controller in {S} and retrieve the predicted robot and control states for that time. Apply the control appropriate for these future states at (possibly future) time M-T.*

5 Predicting Object State

In contrast with the explicit kinematic simulation used to predict the system state, standard optimal (i.e. variance-minimizing) filtering techniques are used to predict the position and velocity of the world object. The simulation has been run with extended Kalman filters, (linear) Kalman filters, and time-invariant filters as predictors. It is standard practice with optimal filtering to use statistical techniques to see if the current dynamic model fits the data, and if not to substitute another model [7,13]. This "variable dimension" approach is the predictive filtering equivalent of the signal synthesis adaptive control scheme [3], and the block diagrams of the two systems are basically the same.

Ideally, pursuit and gaze-shift control can use both retinal and three-dimensional position and velocity data. The retinal frame is in general a poor one in which to express the dynamical properties of exterior objects, since eye movements subject it to complicated time-varying accelerations. Under some conditions, such as smooth pursuit after transients have died out, the object's image may obey a simple (say constant-velocity) dynamical law in retinal coordinates, and application of smoothing or prediction techniques to the retinal image may be appropriate. A more useful two-dimensional coordinate system for some purposes gives the direction of the object in head-centered coordinates which tells where to point the head, and is sufficient for pointing the eyes if the object is distant in relation to the eye baseline. The object's dynamical behavior reduces to its changes of direction in these coordinates, but can still be sufficiently well-behaved to be useful.

For current purposes the choice of filter is unimportant: We use time-invariant filtering appropriate to a LAB-based position sensor, and some control inputs are the spatial coordinates of objects in LAB. Modern frame-rate vision hardware (or sonar) can solve certain versions of the ranging or spatial localization problem in real time. If speed is a problem, the predictive filters can of course accept data at a lower rate than is needed for control output. Although the state of an object (three-dimensional position and velocity) is necessary to do precise pointing of the cameras and head due to perspective and parallax effects, for practical purposes a sufficiently accurate estimate results from direction derived from retinal coordinates and a very approximate range.

5.1 The $\alpha - \beta$ Filter

Linear dynamical systems with *time-invariant* coefficients in their state transition and measurement equations lead to simpler optimal estimation techniques than are needed for the time-varying case. The state estimation covariance and filter gain matrices achieve steady-state values that can often be computed in advance. Two common time-invariant systems are constant-velocity and constant-acceleration systems.

For this work, we assume a constant velocity model: starting with some initial value, the object's velocity in LAB evolves through time by *process noise* of random accelerations, constant during each sampling interval but independent. The cumulative result of the accelerations can in fact change the object's velocity arbitrarily much, so we model a maneuvering object as one with high process noise (Fig. 4). For this work we assume position measurements only are available, subject to measurement noise of constant covariance. Clearly the more that is known *a priori* about the motion the better the predictions will be. Some sensors or techniques can provide retinal or world velocity measurements as well.

Assume the object state (its position and velocity) evolves independently in each of the (X, Y, Z) dimensions. For instance, in the Y dimension, it evolves according to

$$\mathbf{y}(k+1) = \mathbf{F_Y}\mathbf{y}(k) + \mathbf{v}(k), \tag{1}$$

where

$$\mathbf{F_Y} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \tag{2}$$

for sampling interval Δt, and $\mathbf{y} = [Y, \dot{Y}]^T$. The equations for the other two spatial dimensions are similar, and in fact have identical $\mathbf{F}$ matrices. Thus for the complete object state $\mathbf{x} = [X, \dot{X}, Y, \dot{Y}, Z, \dot{Z}]^T$, $\mathbf{F}$ is a (6×6) block-diagonal matrix whose blocks are identical to $\mathbf{F_Y}$. The error vector $\mathbf{v}(k)$ obeys $E(\mathbf{v}(k)\mathbf{v}^T(j)) = \mathbf{Q}\delta_{kj}$.

The $\alpha - \beta$ filter for state prediction has the form

$$\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \begin{bmatrix} \alpha \\ \beta/\Delta t \end{bmatrix} [\mathbf{z}(k+1) - \hat{\mathbf{z}}(k+1|k)], \tag{3}$$

where $\hat{\mathbf{x}}(k+1|k+1)$ is an updated estimate of $\mathbf{x}$ given $\mathbf{z}(k+1)$, the measurement at time $k+1$. Here we assume that $\mathbf{z}(k+1)$ consists of the three state components (X, Y, Z) (but not $(\dot{X}, \dot{Y}, \dot{Z})$). The state estimate is a weighted sum of a state $\hat{\mathbf{x}}(k+1|k)$ *predicted* from the last estimate to be $\mathbf{F}\hat{\mathbf{x}}(k|k)$ and the *innovation*, or difference between a predicted measurement and the actual measurement. The predicted measurement $\hat{\mathbf{z}}(k+1|k)$ is produced by applying (here a trivial) measurement function to the predicted state.

The $\alpha - \beta$ filter is a special case of the Kalman filter. For our assumptions, the optimal values of α and β can be derived (see [7], for example) and depend only on the ratio of the process noise standard deviation and the measurement noise standard deviation. This

* * * Figure 4 about here * * *

Figure 4: *Squares show the velocity of a "constant velocity" system as it evolves in random-walk fashion due to the influence of a white noise process. The continuous line is the velocity estimated by an $\alpha - \beta$ filter. (a) High maneuvering index* ($\lambda = 20.0$) *implies filter trusts its measurements. (b)* $\lambda = 2.0$ *(c)* $\lambda = 0.1$. *Here the initial state estimate and the reliably known process evolution influence the state estimate much more than the measurements.*

ratio is called the object's *maneuvering index* λ, and with the piecewise constant process noise we assume,

$$\alpha = -\frac{\lambda^2 + 8\lambda - (\lambda + 4)\sqrt{\lambda^2 + 8\lambda}}{8} \qquad (4)$$

and

$$\beta = \frac{\lambda^2 + 4\lambda - \lambda\sqrt{\lambda^2 + 8\lambda}}{4}. \qquad (5)$$

The state estimation covariances can be found in closed form as well, and are simple functions of α, β, and the measurement noise standard deviation. Fig. 4 gives a feel for object motions and the performance of the $\alpha - \beta$ filter. An important issue with all Kalman filter variants is that of initialization. Especially if the process noise is assumed small the initialization must be accurate if the filter is to maintain accurate estimates. The steady state filter cannot adjust its gain through time as can the Kalman filter, so covariances must also be time-invariant.

5.2 Estimation and Pursuit Performance

The incorporation of an optimal filter into the control loop was motivated by the necessity of predicting object state for the Smith predictor. We should expect such a filter to have a noticeable and beneficial effect on pursuit performance, and indeed it seems that humans may use stochastic prediction [27,30]. Fig. 5 shows some effects of prediction. In part (a), light squares show the system's pursuit performance (how closely the optical axis is pointed at the object) using only the current noisy image data. Dark squares show system using the output of the $\alpha - \beta$ filter predicting the (X, Y, Z) position of an object evolving with equivalent noise, and generating the predicted image for use by the pursuit system. (An alternative pursuit control uses data in LAB – see Sec. 7.1.) Parts (b) and (c) show "what the system sees" through time: tracking using the filter also "looks" more accurate in the resulting image. In this and all simulation results in the paper, the time axis is in arbitrary units chosen to illustrate the relevant behavior.

* * * Figure 5 about here * * *

Figure 5: *(a) The angular pointing error of the left eye while pursuing a moving object (initially off center in the visual field), measured to the true object position. Light squares: results with current noisy image data. Dark squares: results with output of the $\alpha - \beta$ filter. (b) The position of the object's image on the retina (the x and y data for both left and right eyes is superimposed) when the system is pursuing using current image data as input (the light square case in (a)). (c) As for (b), but when the system is pursuing using the estimated object position as input (the dark square case in (a)).*

* * * Figure 6 about here * * *

Figure 6: *Dark lines give (X, Y, Z) LAB, HEAD, and eye (L and R) coordinate systems. The retinal (x, y) systems coincide with eye X and Y axes. A "cyclopean" system giving direction in spherical (θ, ϕ) coordinates is indicated for the dominant L eye. Straight dashed lines indicate rotation axes for HEAD, pan and tilt axes for eyes. Light solid lines are rigid kinematic links.*

6 Eye and Head Saccade Geometry

6.1 Coordinate Systems

There are four main coordinate systems of interest in this work: LAB, HEAD, and (left and right) eye and retinal. The LAB, HEAD, and eye systems are three-dimensional, right-handed and orthogonal. The retinal system is two-dimensional and orthogonal. LAB is rigidly attached to the environment in which the animate system and objects move. HEAD is rigidly attached to the head. The eye systems are rigidly attached to the cameras, and the retinal systems represent image coordinates resulting from perspective projection of the visible world through the eyes. The camera principal points do not lie on any head rotation, pan, or tilt axis, resulting in parallax effects in imaging and changes in camera position induced by rotations, pans, and tilts (Fig. 6).

The retinal x coordinate is horizontal in the image, increasing rightwards. The vertical y coordinate increases downwards. The eye x and y coordinates parallel the retinal x and y, with eye z pointing out along the camera optical axis, or line of sight. When the eyes are at zero pan and tilt, HEAD x, y, z parallel the eye x, y, z. Let $\mathbf{k}_H$ and $\mathbf{k}_E$ be the unit vectors along HEAD and eye z axes.

6.2 Control Parameters for Saccades

The primary goal of the fast gaze-shift system is to center the image of the object as quickly as possible. Minimizing time means that all necessary degrees of freedom should be adjusted simultaneously.

Errors in x are reduced by panning, in y by tilting. The x and y vectors are not proportional to angle error, since the imaging model is point projection onto a planar, not a spherical, retina. For small angles the error is negligible (for a 28 degree field of view the largest error is about 2%, and for a 57 degree field the largest error is about 10%). If retinal errors are used for pursuit and vergence control the nonlinearity is quickly overcome by feedback. If retinal coordinates are used to generate eye saccades to the periphery, then the above errors can be removed by passing the error signals through an $\arctan(x/f)$ function, where f is the camera's focal length. Retinal data provide information (at least) on object direction, and an (X, Y, Z) position in HEAD or LAB can be constructed consistent with the object's perceived direction and some likely range. According to our assumptions, gaze-shift and pursuit control have available such HEAD or LAB information from sensors or calculations, and hence know or can calculate $\mathbf{O}_H$, the object direction unit vector in HEAD, and $\mathbf{O}_E$, the object's unit direction vector from either eye. If both are available, retinal data can be combined with other sensor data [17]. The assumption of object representations in non-retinal coordinates is significant.

- At close range, spatial information is needed to solve the inverse kinematics necessary to point accurately at the object.
- Range information can speed the necessary reactions. One example is camera focus, and it is interesting that the "near triad" reflex in primates is a coordinated adjustment of vergence, accommodation (focus), and gaze shift when the new object of interest is known to be at a different depth from the old. The near triad speeds eye accommodation: knowing what direction to change avoids "hunting", and the rate of change of focus seems higher as well [28].
- Retinal representations are not robust against eye movements, objects out of view, obscurations, failed image processing, or unexpected plant disturbances. The ability to keep pursuing or to shift gaze correctly despite these surprises is important (Fig. 10(c) in Section 7.3 shows how impressively robust performance in primates can be).
- There is a larger related question of how to represent the visual world. One recent proposal is that objects are represented as object-centered coordinate frames derived from gaze parameters, and linked to each other in a network [4].

6.3 Eye Saccades

Eye saccades are performed by the independent pans and a common tilt. The assumption is that tilt affects the pan axis but not vice-versa. We desire the pan and tilt angles that will point $\mathbf{k}_E$ in the direction $\mathbf{O}_E = (x, y, z)^T$ in eye coordinates. The pan and tilt may be approximated by zero if the head is pointing at the object, or may be calculated from the image or even from a remembered object position and the plant state.

True pans and tilts are derived from an inverse kinematic calculation based on head geometry. For objects distant with respect to the link lengths, approximate eye tilt angle is

$$\theta_{Et} = -\arctan(y/z). \qquad (6)$$

The eye pan angle is

$$\theta_{E_P} = \arcsin(x/R), \qquad (7)$$

where $R = (x^2 + y^2 + z^2)^{1/2}$.

If the maximum angular velocity attainable by the eye motors is $\dot{\theta}_{Emax}$, then this velocity may be simultaneously commanded to pan and tilt for the two required times, or the equivalent number of discrete angular steps commanded. Alternatively the velocities may be scaled so that the pan and tilt are completed at the same time, with the smaller motion performed more slowly. The latter method is used in the simulations. The independence of the tilt axis from the pan motion is all that is needed to allow both motions to proceed simultaneously or in either order with no interference: the situation is different for head saccades.

6.4 Head Saccades

The semantics of HEAD is that of a coordinate system. Even if its axes initially correspond to actual rotation axes, as they can if the origin of HEAD is taken to be the center of a robotic wrist, the situation is not the same as with eye saccades. The control commands for HEAD orientation are to rotate HEAD about one of its axes: such rotation changes both other axes. Since 3-D rotations in general do not commute, the simultaneous rotations we desire will interact with one another. Thus pan and tilt angles computed as for eye saccades will thus only work correctly if the movements are carried out sequentially in the correct order.

Since simultaneous rotations are needed for speed, we must compute the continuous rotation velocities about HEAD Y and Y axes needed to accomplish a head saccade, and the time interval to activate them. For discrete simulation or implementation as a set of individual commands, we need to approximate the motion by a number N of small rotations. Depending on the accuracy needed, a number of these small rotations may be composed to correspond to what happens in one simulation instant. Luckily, small rotations are the key to the non-commutativity problem, and the error in the approximation can be made to fall off as N^{-2}.

Consider a smooth rotation taking $\mathbf{k}_H$ to $\mathbf{O}_H = (x, y, z)^T$ directly (along a great circle route on the HEAD-centered Gaussian sphere) and at a constant speed. The angular distance to be covered is

$$\theta_{tot} = \arccos(\mathbf{k}_H \cdot \mathbf{O}_H) = \arccos(z). \qquad (8)$$

Projecting the Gaussian sphere onto the (X, Y) plane projects the great circle along which $\mathbf{O}_H$ moves onto a straight line from the origin to (x, y). Thus the instantaneous direction of the head motion is constant, and at any time the ratio of distance rotated around x and y axes is $x/(-y)$. If the total angular distance to be covered is divided into N steps, each step consists of small head rotations θ_{Hx} and θ_{Hy}, where $\theta_{Hx} = (x/(-y))\theta_{Hy}$. A Euclidean approximation yielding the desired distance θ_{tot} is that

$$\theta_{Hx} = (x(\theta_{tot}/N))/r \qquad (9)$$

* * * Figure 7 about here * * *

Figure 7: *Hollow squares show the $1/N$ angular cumulative error (radians) of a sequence of [pan, tilt] motions approximating a particular smooth motion with simultaneous pan and tilt. Dark squares show th $1/N^2$ error for a sequence of [pan, tilt] [tilt, pan] motions. The abscissa is the number of [pan, tilt] or [tilt,pan] component motions in the sequence.*

and

$$\theta_{Hy} = -(y(\theta_{tot}/N))/r, \tag{10}$$

where $r = (x^2+y^2)^{1/2}$. The error increases as the spherical triangle differs from the planar. If $\dot{\theta}_{Hmax}$ is the maximum rotational velocity of the head motors, the continuous time to complete the entire head saccade is

$$t = (\max(x,y)(\theta_{tot}))/(r\dot{\theta}_{max}), \tag{11}$$

with the velocities in the ratio x/y, so assigned that the pans and tilts complete at the same time and the greater velocity is $\dot{\theta}_{max}$. The approximation of the continuous motion by a sequence of small pans and tilts works because infinitesimal rotations about coordinate axes, as the generators of the Lie group of rotations, commute [1]. The resulting cumulative angular error decreases reliably as $1/N$, independent of z or x/y (except that the error is zero if either of x or y is zero). If the desired direction is $(x,y,z)^T$, the result for the worst case, $N = 1$, is obtained by applying the tilt and pan given by eqs. (9) (10) to the vector $(0,0,1)^T$. The resulting vector is

$$\hat{\mathrm{O}} = [\cos(\theta_{Hy})\sin(\theta_{Hx}), \sin(\theta_{Hy}), \cos(\theta_{Hy})\cos(\theta_{Hx})]^T. \tag{12}$$

The difference between the desired vector and the result ($\|\hat{\mathrm{O}} - \mathrm{O}_H\|$ or $\arccos(\hat{\mathrm{O}} \cdot \mathrm{O}_H)$) may be scaled by $1/N$ to predict the corresponding error in the cumulative motion. If the simulation or discrete command sequence uses alternating [pan, tilt] and [tilt, pan] actions in the sequence of small motions, the error decreases as $1/N^2$. Fig. 7 shows the comparison for a large head saccade "back over the right shoulder".

7 Eye-Head Saccades

There are two main reasons for head motions during gaze shifts. First, they can keep the eyes centered in the head, maximizing the range of eye rotation before encountering mechanical stops. Second, they can increase the velocity of eye rotation in LAB, leading to a more rapid visual acquisition of the desired object.

Saccadic eye and head motion in primates was for a long time believed to proceed without visual or proprioceptive feedback, and there are recent models of head saccades using three-burst bang-bang control [39]. However, there is also strong evidence that the eye's saccadic system can correct its commanded motions after the retinal stimulus is removed, but before the motion has begun, to deal with certain disturbances applied to

* * * Figure 8 about here * * *

Figure 8: *Linear summation. Robot gaze (butterflies), eye (light squares), and head (dark squares) angles showing linear summation effect arising from reflex interaction. Here the eye saccade control is active together with with head-compensation and VOR reflexes. The combination of eye saccades with head compensation and VOR forces the gaze velocity to be the velocity of the eye saccade acting by itself. Time units are arbitrary and chosen to illustrate the behavior.*

the eyes (e.g. [25]). This evidence is consistent with the "local feedback" hypothesis [34] that the eye saccadic system is guided by an internal representation of its effects in LAB or HEAD coordinates rather than by retinal error. The basic idea of relating eye commands to the world rather than to the retina of course extends beyond saccadic commands (for pursuit, see [42]). The idea is entirely consistent with the modern control theory ideas of state feedback using state observers (e.g. [19]), and with predictive techniques in gaze control (e.g. [11]).

This section describes four versions of cooperating eye-head motions. In each, the goal of the head saccadic system is taken to be *position matching*: When the head saccade finishes the head is pointing at the (instantaneous) current object position and head rotation velocity commands are zero. The goal of the eye saccadic system is taken to be *position and velocity matching*. It would be just as easy to produce position and velocity matching in both systems, to move the head by only the minimum amount that allows acquisition, or other variations.

The goal is by no means to reproduce the complex findings on primate systems and their interaction [33,21], but to establish that a usable range of capabilities is possible. Although certain primate capabilities (such as coping with disturbances) have inspired features of some of the versions, the goal has simply been to produce *fast, stable, and robust* eye-head gaze shifts. The first two schemes have the virtue of simplicity, but are handicapped by control interactions and use no prediction. The third scheme has no feedback and optimal speed, but no robustness. The last and most successful scheme uses explicit predictive simulation to generate gaze shifts, and the loop is closed by re-triggering the calculation when a disturbance is sensed.

A first example of an eye-head saccade is provided by simply running the "sampled" eye saccades as defined in Section 3.1) simultaneously with the "continuous" VOR and head-compensation reflexes ([11], Section 3.1). Fig. 8 shows the result, which is similar to that which once was claimed (under the "linear summation hypothesis") to hold in primates, *viz.* that their eye rotational velocity measured in LAB is kept constant (eye saccadic velocity is decreased by head saccadic velocity) [29].

In primates, the linear summation hypothesis has been contradicted by findings indicating eye-head gaze shifts outperform eye saccades alone [36] (see Fig. 9(a)). Certainly in a robotic context a speed improvement is attainable, and the following three versions are intended to outperform the first one. The *Gaze Feedback* version is inspired by local feedback theories, and is based on "continuous" head and eye position-feedback controls.

A head saccade is then simply implemented as a pursuit operation, and cooperating eye-head gaze-shifts are simultaneous head and eye pursuit. The *Sampled Optimal* version is a closed-form "sampled" solution for full-speed eye and head movements. The *Simulated Optimal* version uses the familiar pipeline of predicted future states to maintain optimal speed and add the "continuous" ability to deal with varying velocities and disturbances.

7.1 Gaze Feedback

This version of eye-head saccades is just the simultaneous action of pursuit feedback loops in both eye and head. Thus in this section the "sampled", maximum speed eye control is replaced by a "continuous" pursuit control and the head is treated similarly. The same controls may thus be used for saccades as for pursuit. This literal and straightforward interpretation of "local feedback" affords some illuminating comparisons.

In one obvious local feedback scheme the eye pursues its desired position with respect to the head. This means that its final position in HEAD is entirely determined by the range of the object and head geometry. Given that the system knows the range R of the object and the horizontal and vertical offsets (h and v) of the eyes from the head origin, the eye position when k_H (the head) is pointed at the object is $(pan, tilt) = (\arctan(h/R), -\arctan(v/R))$. Being at this position only centers the object when the head is correctly pointed, and so using this target for pursuit limits the eye-head saccade to the speed of the head saccade alone, even if the eye achieves its final position instantaneously.

An equally obvious alternative is to have eye and head both pursue the object position (in LAB or retinal coordinates). This scheme would seem capable of faster performance, but in practice is disappointing. Feedback lowers gains, and PID control does not cope with maximum velocities (saturation) gracefully. Second, control interactions change the system to fourth-order, with a performance penalty.

Fig. 9 shows two comparable simulated eye-head "saccades" implemented with position error feedback, along with human data from [36]. Incidentally, this instance demonstrates the LAB-based pursuit system: Its error signal is the angular difference (in pan and tilt directions) between the camera axis and the direction of the object. No significant differences arise when the eyes pursue the retinal image if the object is distant. In the human (Fig. 9(a)), gaze converges quickly to the object, and head motion speeds gaze shift completion by a factor of two. In Fig. 9(b), VOR is active. Its effect is to slow the eye rotation by the amount of head rotation, and thus the behavior of the gaze angle is the same as if the eye "saccade" (now pursuit) system were acting alone. Just as in the linear summation gaze shift above, the VOR again defeats the goal of moving the gaze to the object quicker than the eye saccade acting alone.

In the case of Fig. 9(c), the eye-head saccade is made with VOR turned off. The head is pursuing the object as a second order system, carrying the eye with it. The eye control is also second order, but is being forced by the head's motions. The effect is that of a fourth order system, and the case of a saccade to a motionless target corresponds to a step

* * * Figure 9 about here * * *

Figure 9: *Gaze feedback. (a) Human gaze (G), eye (E), and head (H) movements during 80 degree gaze shift. A has no head motion, B has 80 degree head movement, and gaze shift is completed twice as fast. (b) Robot gaze (butterflies), eye (light squares), and head (dark squares) angles for gaze feedback saccade. Both head and eyes move to the object position under the influence of a smooth pursuit PID feedback control law driven by position error. VOR is active, so eye movement is slowed and performance is identical to an eye-only saccade. (c) As for (b), but VOR is not active. The resulting fourth-order system has no better performance than in (b).*

input. The root locus behavior of fourth order systems makes them prone to instability as gain increases. Their performance is also at issue. Optimal coefficients for nth-order systems under the several performance criteria (such as the integral of time multiplied by absolute error) may be derived, and step responses for the optimal systems of various orders plotted [16]. Two points emerge.

1. Under three common performance criteria, optimal fourth-order rise times are slower than optimal second order, and convergence to the desired value is no faster.
2. The cascaded second-order systems give a restricted form of the general fourth order system that does not in general produce the optimal higher-order system, and certainly does not if the lower order systems are themselves optimal.

These observations are borne out by experimentation with the PID gains for the head and eye pursuit. Since head-eye rise time will be faster than either system acting alone (unless they rotate in opposite directions), the resulting system is guaranteed to be sub-optimal by point 1 above, and even the optimal fourth-order system could perform no better than either system acting alone.

To summarize: Straightforward implementations of gaze feedback models have performance limitations. If eye pursuit moves to a position in HEAD the gaze shift of a head-eye saccade cannot be faster than the head saccade by definition. If a VOR-like compensation is in effect, a head-eye saccade cannot be faster than the eye saccade by definition. Last, using control without compensation, the combined gaze motion obeys a higher-order form of control whose performance may well be theoretically worse than either eye or head acting alone. The psychophysical data indicate that the primate system exceeds the speed of either head or eye saccades without oscillations. Thus gaze feedback seems to have technical problems that make it unsatisfactory for use with PID control and perhaps unpromising as a model for primate abilities as well. The next versions incorporate time-optimal ("bang bang") control.

7.2 Sampled Optimal

Time optimality demands that the eye and head both move at maximum velocity to acquire the (possibly moving) object. For generality we assume arbitrary head and eye initial positions H_0 and E_0, along with the initial object position T_0. We assume a control delay Δ_H for head rotations, and a control delay Δ_E for eye rotations. (Maximum) head velocity is v_H, (maximum) eye velocity is v_E, object angular velocity in HEAD is v_T.

The commands for head and eye saccades are assumed to be computed instantaneously and issued at time 0. (Thus they begin at times Δ_H and Δ_E). The directions for the saccades are easily computed in each dimension. The problem is to compute the durations of the saccades, or equivalently the times the velocity commands should cease, t_H and t_E.

The algorithm computes the head saccade first, since it affects eye position and not vice-versa. At t_H the object position and the head position are to be equal, so (in one dimension)

$$T_0 + t_H v_T = H_0 + (t_H - \Delta_H)v_H. \tag{13}$$

thus

$$t_H = \frac{T_0 - H_0 + \Delta_H v_H}{v_H - v_T}. \tag{14}$$

Computing t_E is more complicated. There are six cases, depending on the overlap between the periods of head and eye motion. The cases can be easily distinguished by simple tests based on the delays, velocities, and t_H. For example, the computation for the case in which the eye departs after the head and also reaches its final position after the head does,

$$T_0 + t_E v_T = E_0 + (t_H - \Delta_E)(v_H + v_E) + (t_E - t_H)v_E \tag{15}$$

and so

$$t_E = \frac{T_0 - E_0 - t_H v_H + \Delta_E(v_H + v_E)}{v_E - v_T}. \tag{16}$$

The expression for t_H and the six expressions for t_E all have the form

$$t_X = f(T_0, v_T, X_0; c_1, c_2) = \frac{T_0 - X_0 + c_1}{c_2 - v_T}, \tag{17}$$

where X stands for E or H, and c_1 and c_2 depend on the case but are simple combinations of the (presumably unchanging) eye and head delays and maximum velocities. The expressions for t_X become linear when the object is stationary ($v_T = 0$).

This version gives optimal performance if the various assumptions are met. Implementing it with simulation and recalculating when necessary yields the next method, which is more robust.

7.3 Simulated Optimal

This version is basically an implementation of the last method using the predictive simulation techniques of the main gaze control system. The saccade control simulation uses the pipeline of future system states. The important quantities are the angular position

* * * Figure 10 about here * * *

Figure 10: *(a) Eye (x, y) position, and (b) eye (dark squares) and head (light squares) x-velocity for eye-head saccade to moving object. Head saccade starts at $k = 3$, adding its velocity to eye saccade. Eye reaches object at $k = 9$, head at $k = 21$. (c) Data from primate saccades with eye velocities disturbed (eyes dragged down to left) prior to saccade but after stimulus has vanished. (d,e) As in (a,b), showing eye-only saccade with disturbance during $3 \leq k \leq 10$ adding negative (x, y) velocity before and during first part of saccade. (f,g) As in (a,b), with eye-head saccade and disturbance during $3 \leq k \leq 7$.*

and velocity for the object (in HEAD coordinates) head (in LAB), and eye (in HEAD), indexed by time: $\mathbf{O}(k), \dot{\mathbf{O}}(k), \mathbf{H}(k), \dot{\mathbf{H}}(k), \mathbf{E}(k), \dot{\mathbf{E}}(k)$.

The simulation algorithm first computes a head saccade by determining, for the x and y dimensions, the direction to rotate and then driving at maximum velocity, stopping after there is a zero-crossing in predicted object position (say at time k). The head position at time $k - 1$ may be closer to the object: the closest position is chosen. For more accuracy, time k could be used and (assuming locally constant object velocity) the head motion for the last interval slowed to

$$s = \dot{\theta}_{max} \frac{\mathbf{O}(k-1)}{\mathbf{O}(k-1) - \mathbf{O}(k)}. \tag{18}$$

The sequence of head-saccade commands is inserted into the pipeline and the resulting changes to head and eye position computed (these values overwrite existing values, as the saccade is taken to replace previous control). Then the eye saccade is computed exactly like the head saccade, except eye position is computed by adding the head-relative eye velocity to the head velocity.

So far, this version of saccade control is "sampled", and in simple cases the solution obtained is the same as the sampled optimal solution above. However the simulation solution copes uniformly with head and eye saccade interactions, and handles arbitrary head or eye velocity profiles (velocity ramp-ups and ramp-downs, for instance) as well as arbitrary (predictable) object motion. To cope with disturbances, and to implement successive ("catch up") saccades, the simulation is placed in a loop that runs until the saccade is successfully completed, checking if at any time a new calculation must be performed. In the simulation, a disturbance to the velocity or position of the eyes is taken to signal the need for a recalculation. This gives the robustness of a "continuous" system only computed on demand. Fig. 10 shows some results from this control scheme. After reaching the object, the eye continues moving at the object's velocity, (due to smooth pursuit or velocity matching by the eye saccade), and VOR compensates for any continuing head motion. When macaque monkeys have their eye velocities electrically disturbed after the target stimulus has vanished, they successfully correct the saccade (Fig. 10(c)). The recalculation can sometimes be done during the latency period of the original saccade, so the correction does not delay saccade onset [25]. In the simulated system the control delay is unavoidable, but correction does occur (Fig. 10(d-g)).

8 Discussion

Prediction is the key to stability of interacting closed-loop control systems with time delays. Gaze control is one of several central capabilities needed by a behaving sensorimotor system to support higher-level activities. Both biological and robotic systems have delays and interactions making simple feedback schemes inadequate for gaze control. This work demonstrates kinematic and stochastic models for prediction in a simulated system with multiple interacting gaze control "reflexes" implemented as various hybrids of "continuous", "sampled", and predictive control. Currently there are ten reflexes, including the velocity-error version of smooth pursuit and both retinal and spatial smooth pursuit and saccadic systems, with four operating simultaneously in pursuit, and between two and four in versions of saccades. Earlier work [11] demonstrated the viability of Smith prediction for simple gaze control. This paper extends the prediction to include the behavior of the object of attention in three dimensions, and extends the gaze-shift capabilities to include cooperating head and eye saccades that are robust against sensed plant disturbances. Aspects of primate gaze-control inspired certain techniques, but no effort has been made to model or duplicate quantitatively any biological capability *per se* (for more on this topic see [11]). Similarities sometimes arise, probably through basic agreement as to the necessities for a practical foundation for visual perception in a behaving agent.

If a gaze control system like the simulated one is to become reality, significant further work on hardware and on systems and application software is needed. Leaving these large issues aside, there is much work to be done in implementing the controls. A first step is to parameterize the simulation with time constants, velocities, and variances that reflect the true robotic situation. High variance in time delays from the current software may make cooperation between controls impossible: the applications programmed so far perform successfully, but only with one feedback loop in operation. We anticipate that more sophisticated hardware and software will soon make delays more predictable. Empirical and theoretical sensitivity analysis of the system leads to parameter-adaptive control that can compensate for slowly-varying plant parameters or delays [24]. Computation time is probably not a difficulty: the basic kinematic simulation including the predicted image requires fourteen 3×3 matrix products (six of which involve sparse and uniformly-structured matrices), six cosine calculations, six square roots, and two vector additions. The $\alpha - \beta$ filter needs two more matrix products, two matrix additions, and about 15 scalar multiplies for a total of about 350 floating point multiplies and 250 additions. This expense must be borne for every position in the pipeline, and ignores a certain amount of other data management. The controls can to some extent run in parallel (although there is a shared data structure.) The simulation can be speeded up by discretizing at a larger time interval. Multiple Kalman filters may be needed instead of the $\alpha - \beta$ filter to deal with HEAD-based position measurement techniques, multiple objects, and objects of differing dynamic characteristics. The PID controls must be modified to deal with saturation. Sensor fusion techniques [17] can be used to combine the sensor outputs to obtain more accurate position information. A cyclopean image frame may be needed in which to integrate left and right images for disparity calculations and other cooperative image analysis tasks. The obvious choice is that of the dominant eye, but instead of cartesion coordinates, some version of spherical projection may be worth the effort for the invariance it provides [22]. The interaction of gaze control with attentional, reflexive, and planning processes remains a live issue that becomes even more interesting with both foveal and peripheral visual capabilities.

The hope was that a gaze control module could be designed that would manage the cameras, present a simple interface to the higher cognitive levels, and carry out appropriate behavior autonomously: gaze control would be one level in a "subsumption"-like hierarchical architecture [10]. It seems that some approximation to this goal is possible using a range of control techniques acting together using predictive techniques, but also that other powerful paradigms, forms of adaptation and learning, are also at work. The psychophysical data, and the experience with the simulator, suggest that in primates gaze control could itself be described as a problem-solving, goal-directed activity, in which various lower-level capabilities are mixed and matched to achieve a desired effect. Most visual "reflexes" seem to a great degree cognitively mutable, and to display various forms of sophisticated adaptive and interactive behavior [21,8].

Control is a good metaphor for management of gaze parameters, and control theory provides a powerful framework and formalism to describe many phenomena in complex systems. Most aspects of gaze control can, under reductionist circumstances, be adequately described as practically implementable control processes. Another metaphor for the gaze control system may be that of a set of *skills* deployed for practical (sometimes including expressive) purposes.

> ...investigations of skill speak not of responses to stimuli but of "strategies" and "procedures" – larger organizations which, although less precisely defined, are closer to describing the units of performance that seem to occur in normal, everyday life. [41]

Though classical motor skill research does not have much in common with research in basic sensorimotor systems such as walking or gaze control [18,41], control theory is a central metaphor or model in both. Motor skills have much in common with the behavior of the gaze "control" system. Parallels exist at several levels. Both performance in a skill and the gain of VOR or OKR vary with attention or motivation. Like saccades, some skilled action commands can be modified during their latency period without slowing onset time (Fig. 10(c) from [25], for hand motions see [26]). Other similarities occur in the tuning of individual components and their orchestrated interaction, basically for the purpose of overcoming delays, maximizing some performance index, and dealing with disturbances (Recall the "near triad" from Section 6.2, and everyday motor skills).

Viewing gaze control as a set of skills provides a different way to structure its investigation and synthesis. The robot gaze control designer, rather than mimicking a particular biological system, can consider the intended task domain to identify and specify functionally important capabilities. So far the simulations have addressed vergence, stabilization, compensation, pursuit and rapid gaze shifts, which are of general use and functionally important in fixating objects in relative motion and redirecting visual resources. However, there are many other tasks an animate system will be engaged in that will call for other visual skills. One is *exploration*: in this context an appropriate reflex might override the VOR and redirect the eyes in the direction the head is turning, thus anticipating the need for vision. As it happens, this particular behavior is found in primates in the so-called "quick phase" eye motions. A basic task is *object avoidance*, while navigating or passive. Primates have the "looming" reflex and other skills, perhaps using optic flow input, like blinking.

In important issue is the use of more sophisticated visual processing as input for more complex reflexes, and some work is being done on this topic already [31]. Each robotic system will have its own idiosyncrasies and strengths, suggesting idiosyncratic skills. For instance, many two-camera robotic heads have a single vergence degree of freedom instead of two independent pan axes. Clearly the details of saccadic, pursuit, and vergence skills must account for this difference. A head with independent pan degrees of freedom like the RR's can be endowed with skills in which the eyes operate more independently, either with shared or independent goals. In the RR's case the shared tilt platform means that any such skills must be coordinated to deal with shared resources, but this is a general problem and indeed is part of the definition of skill.

Control theory is evolving ever more sophisticated methods for incorporating adaptive techniques leading to synthesis of optimal performance applicable to motor learning in general and hence gaze control [3,20]. Though certain advanced control techniques are still impossible to implement quickly enough for robotics, control theory seems increasingly relevant to what might be called skill acquisition. Gaze control is obviously amenable to investigation by computer- and neuroscientists, and there is a wealth of accumulated data and knowledge. If indeed gaze control skills lie somewhere between purely reflexive (innate) and skilled (learned or practiced) behavior, then perhaps they are a promising domain for the investigation of control theory, algorithms and neural mechanisms underlying the acquisition and performance of general sensorimotor skills.

9 Bibliography

[1] S. L. Altmann. *Rotations, Quaternions, and Double Groups.* Clarendon Press, Oxford, 1986.

[2] R. L. Andersson. *A Robot Ping-Pong Player: Experiment in Real-Time Intelligent Control.* MIT Press, 1988.

[3] A. T. Bahill and J. D. McDonald. Adaptive control models for saccadic and smooth pursuit eye movements. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[4] D. H. Ballard. Animate vision. In *International Joint Conference on AI-89*, August 1989.

[5] D. H. Ballard. Behavioural constraints on animate vision. *Image and Vision Computing*, 7(1):3–9, 1989.

[6] D. H. Ballard and A. Ozcandarli. Real-time kinetic depth. In *Second Int. Conf. on Computer Vision*, November 1988.

[7] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1988.

[8] A. Berthoz and G. Melvill Jones. *Adaptive Mechanisms in Gaze Control: Facts and Theories*. Elsevier, 1985.

[9] B. Biguer and C. Prablanc. Modulation of the vestibulo-ocular reflex in eye-head orientation as a function of target distance in man. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[10] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA-2:14–23, 1986.

[11] C. M. Brown. Gaze controls with interactions and delays. In *DARPA IU Workshop. Submitted IEEE-TSMC*, 1989.

[12] C. M. Brown. *The Rochester Robot*. Technical Report 257, University of Rochester, September 1988.

[13] C. M. Brown, H. Durrant-Whyte, J. Leonard, and B. Rao. Centralized and noncentralized kalman filtering techniques for tracking and control. In *DARPA IU Workshop*, 1989.

[14] A. Buizza, R. Schmid, and J. Droulez. Influence of linear acceleration on oculomotor control. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[15] J. J. Clark and N. J. Ferrier. Modal control of an attentive vision system. In *Second Int. Joint Conference on Computer Vision*, November 1988.

[16] Richard C. Dorf. *Modern Control Systems*. Addison Wesley, 1986.

[17] J. M. Brady (Ed.). Special issue on sensor data fusion. *International Journal of Robotics Research*, 7(6), 1988.

[18] K. Connolly (Ed.). *Mechanisms of Motor Skill Development*. Academic Press, 1970.

[19] Pieter Eykhoff. *System Identification: Parameter and State Estimation*. Wiley and Sons, 1974.

[20] H. Flashner, A. Beuter, and C. Boettger. Parameter optimization model of learning in stepping motion. *Biological Cybernetics*, 60:277–284, 1989.

[21] A. F. Fuchs and W. Becker. *Progress in Oculomotor Research*. Elsevier, 1981.

[22] K. Kanatani. Camera rotation invariance of image characteristics. *Computer Vision, Graphics and Image Processing*, 39:328–354, 1987.

[23] A. Kawato, K. Furukawa, and R. Suzuki. A hierarchical neural network model for control and learning of voluntary movememt. *Biological Cybernetics*, 57:169–185, 1987.

[24] J. E. Marshall. *Control of Time-Delay Systems.* Peter Peregrinus Ltd., 1979.

[25] L. Mays and D. Sparks. The localization of saccade targets using a combination of retinal and eye position information. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[26] E. D. Megaw. Possible modification to a rapid ongoing programmed manual response. *Brain Research*, 71:425–441, 1974.

[27] J. A. Michael and G. Melvill Jones. Dependence of visual tracking capability upon stimulus predictability. *Vision Research*, 6:707–716, 1966.

[28] F. A. Miles. Adaptive regulation in the vergence and accommodation control systems. In A. Berthoz and G. Melvill Jones, editors, *Adaptive Mechanisms in Gaze Control*, Elsevier, 1985.

[29] P. Morasso, E. Bizzi, and J. Dichgans. Adjustment of saccade characteristics during head movements. *Experimental Brain Research*, 16:492–500, 1973.

[30] P. D. Neilson, N. J. O'Dwyer, and M. D. Neilson. Stochastic prediction in pursuit tracking. *Biological Cybernetics*, 58:113–122, 1988.

[31] R. A. Nelson and Y. Aloimonos. Using flow field divergence for obstacle avoidance: towards qualitative vision. In *Int. Conf. on Computer Vision 2*, December 1988.

[32] T. Olson and R. Potter. Real-time vergence control. In *Computer Vision and Pattern Recognition 1989*, June 1989.

[33] B. W. Peterson and F. J. Richmond. *Control of Head Movement.* Oxford University Press, 1988.

[34] D. A. Robinson. Oculomotor control signals. In G. Lennerstrand and P. Bach-y-Rita, editors, *Basic Mechanisms of Ocular Motility and Their Clinical Implications*, Oxford: Pergamon, 1975.

[35] D. A. Robinson. Why visuomotor systems don't like negative feedback and how they avoid it. In M. A. Arbib and A. R. Hanson, editors, *Vision, Brain, and Cooperative Computation*, MIT Press, 1988.

[36] D. A. Robinson and D. S. Zee. Theoretical considerations of the function and circuitry of various rapid eye movements. In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[37] O. J. M. Smith. Closer control of loops with dead time. *Chemical Engg. Prog. Trans.*, 53(5):217–219, 1957.

[38] O. J. M. Smith. *Feedback Control Systems.* McGraw-Hill, 1958.

[39] L. Stark, W. H. Zangemeister, and B. Hannaford. Head movement models, optimal control theory, and clinical application. In B. W. Peterson and F. J. Richmond, editors, *Control of Head Movement*, Oxford University Press, 1988.

[40] R. D. Tomlinson and D. A. Robinson. Is the vestibulo-ocular reflex cancelled by smooth pursuit? In A. F. Fuchs and W. Becker, editors, *Progress in Oculomotor Research*, Elsevier, 1981.

[41] A. T. Welford and L. E. Bourne. *Skilled Performance: Perceptual and Motor Skills.* Scott, Foresman, and Co., 1976.

[42] L. R. Young. Pursuit eye movement – what is being pursued? *Dev. Neurosci.: Control of Gaze by Brain Stem Neurons*, 1:29–36, 1977.

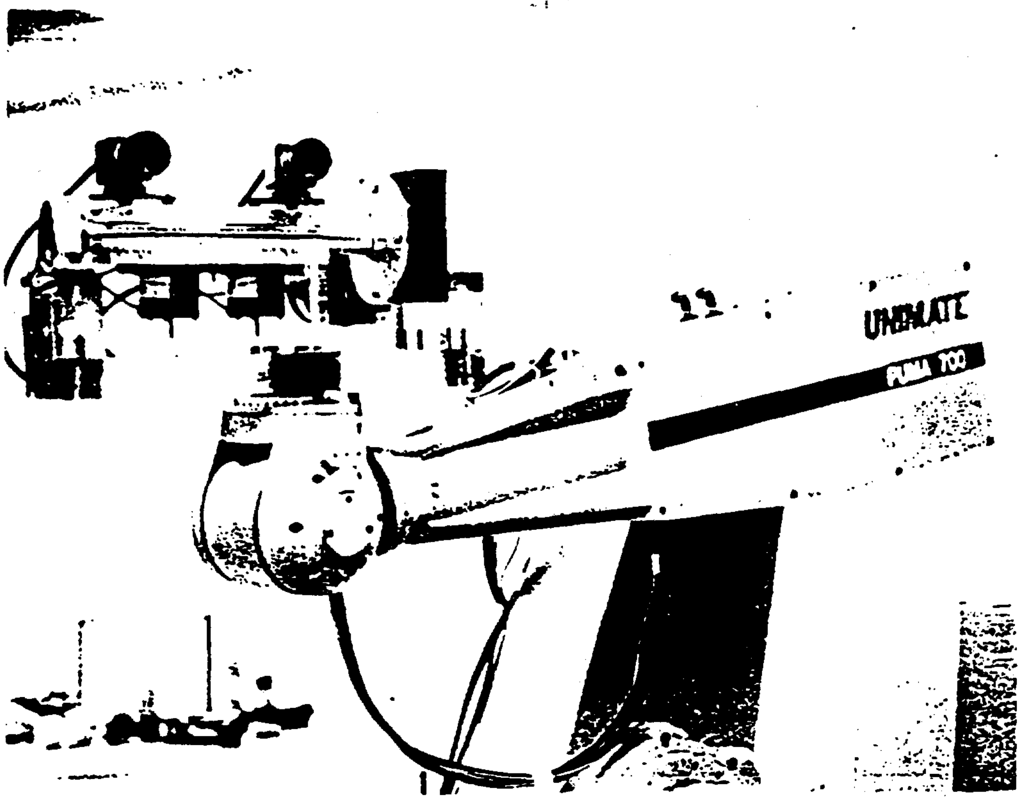

Figure 1

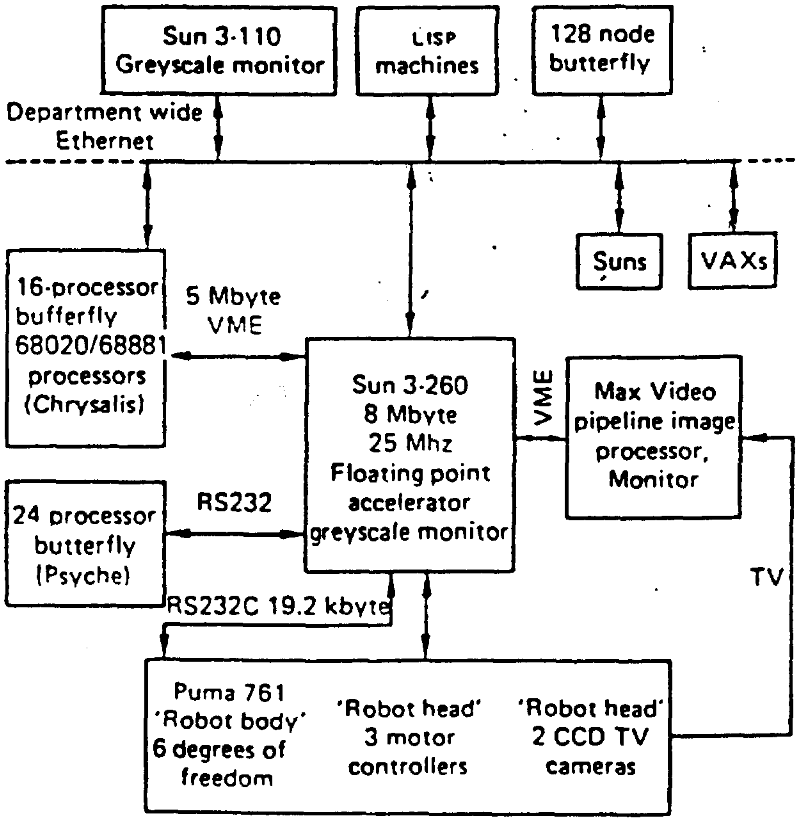
Sun 3-110 Greyscale monitor
LISP machines
128 node butterfly
Department wide Ethernet
Suns
VAXs
16-processor bufferfly 68020/68881 processors (Chrysalis)
5 Mbyte VME
Sun 3-260 8 Mbyte 25 Mhz Floating point accelerator greyscale monitor
VME
Max Video pipeline image processor, Monitor
24 processor butterfly (Psyche)
RS232
RS232C 19.2 kbyte
TV
Puma 761 'Robot body' 6 degrees of freedom
'Robot head' 3 motor controllers
'Robot head' 2 CCD TV cameras

Figure 2

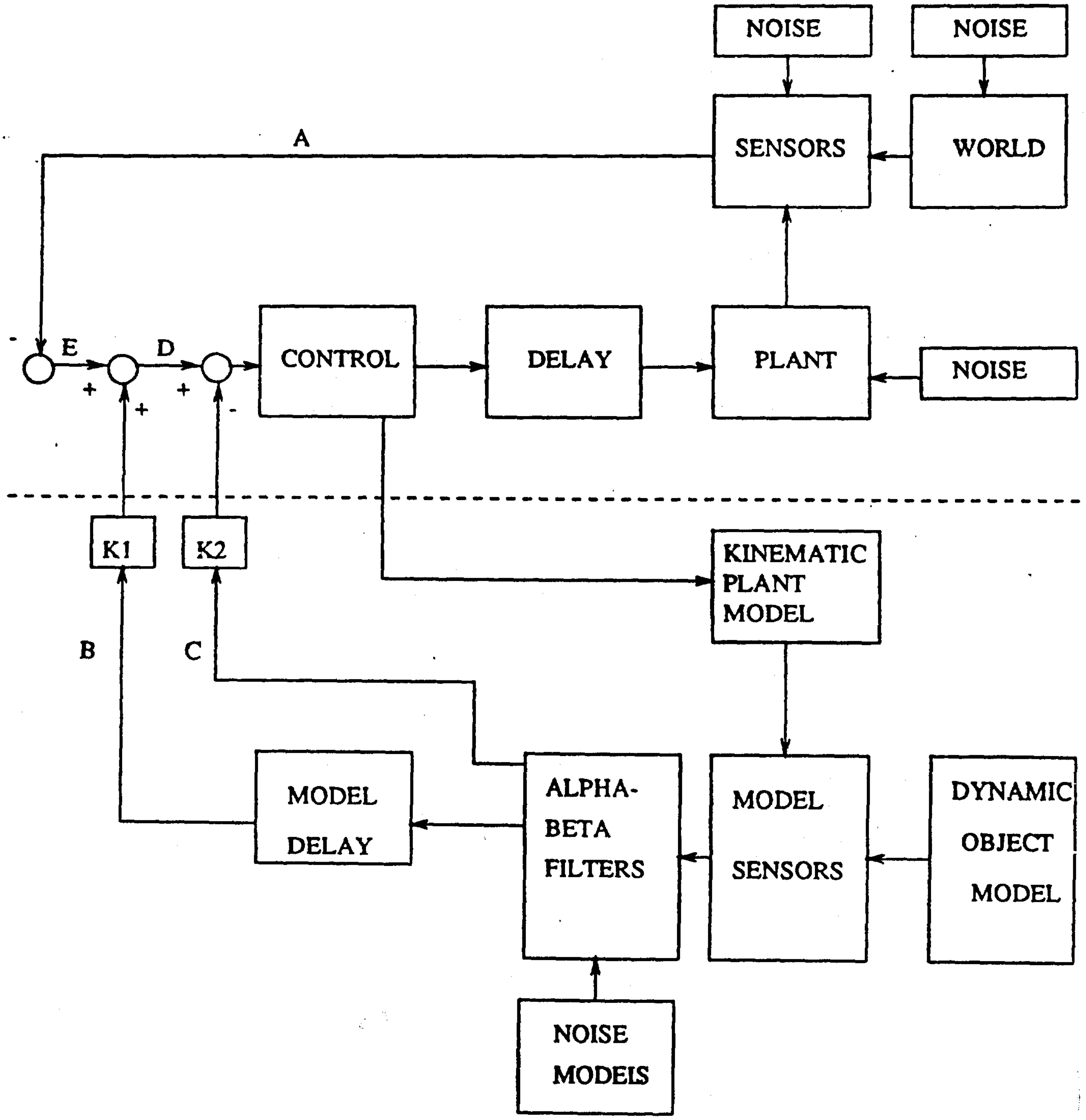
NOISE
NOISE
A
SENSORS
WORLD
-
E
D
+
+
+
-
CONTROL
DELAY
PLANT
NOISE
K1
K2
KINEMATIC PLANT MODEL
B
C
MODEL DELAY
ALPHA-BETA FILTERS
MODEL SENSORS
DYNAMIC OBJECT MODEL
NOISE MODEIS

Figure 3

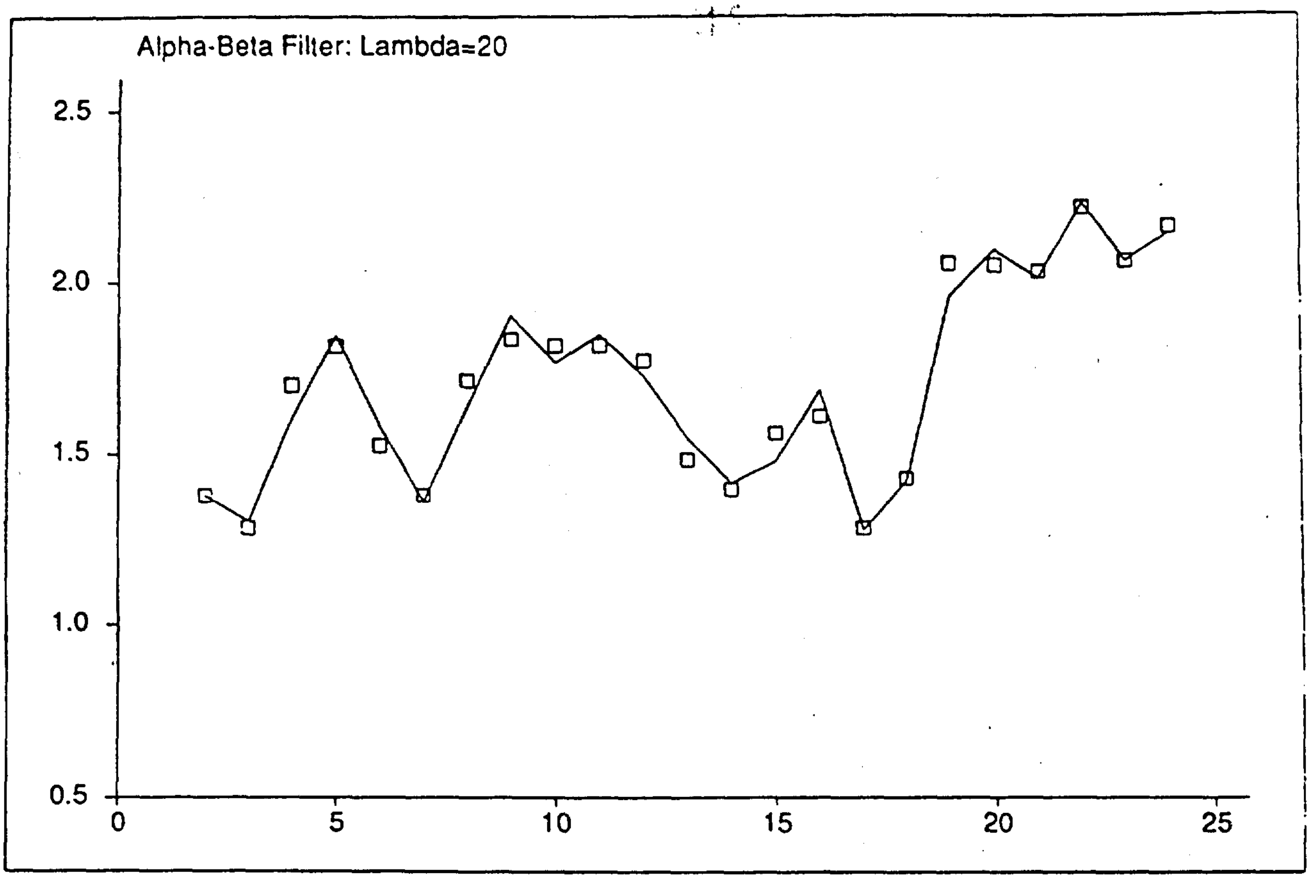

Figure 4a

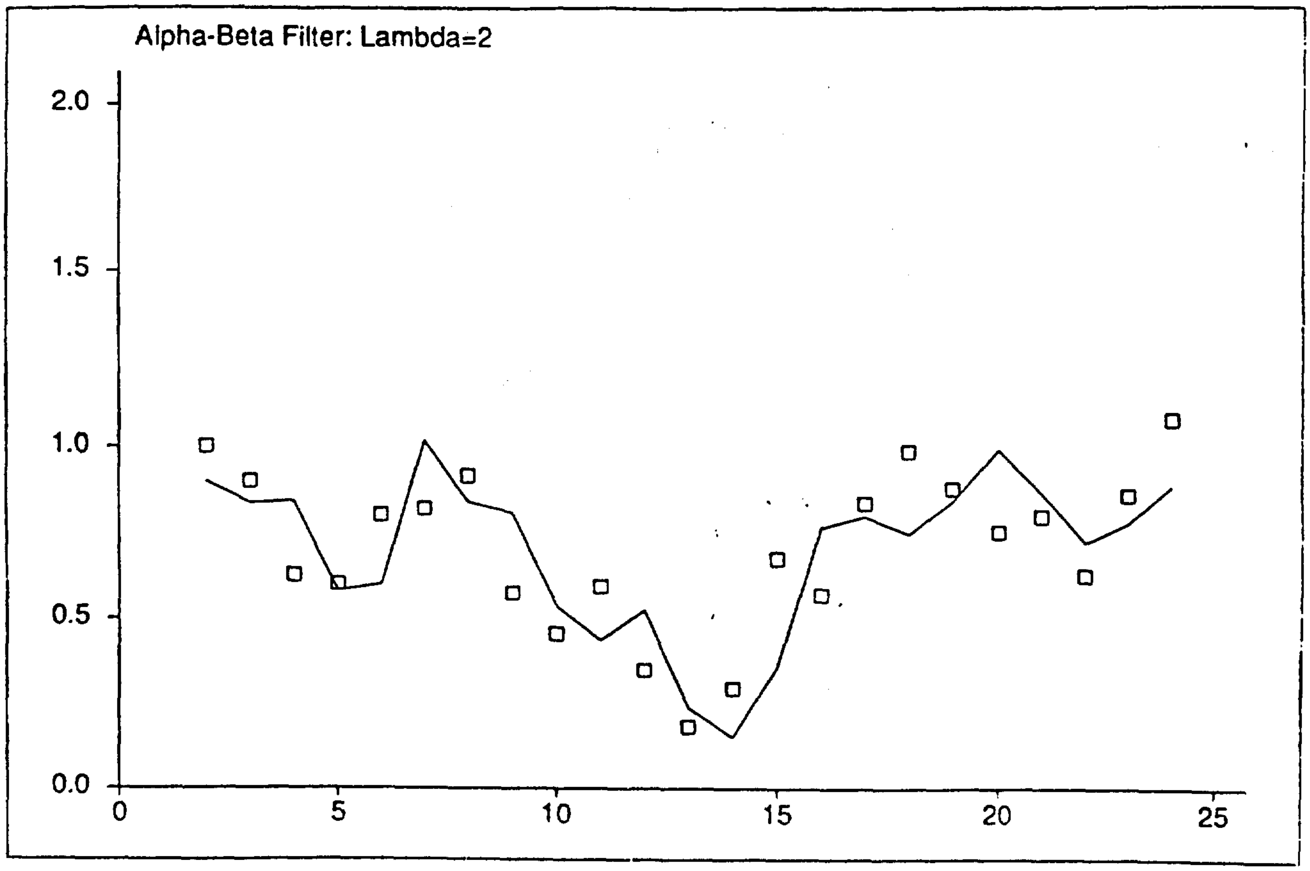

Figure 4b

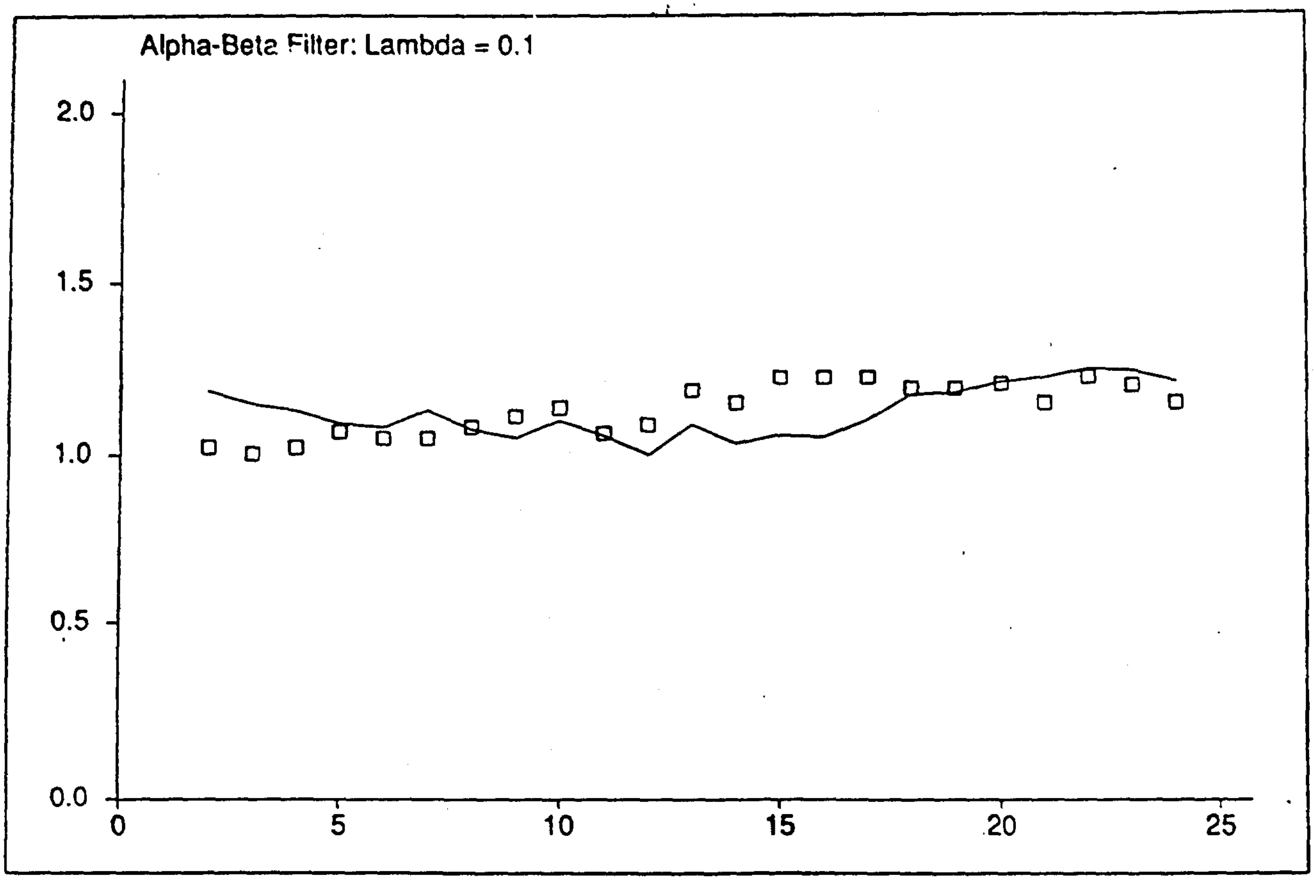

Figure 4c

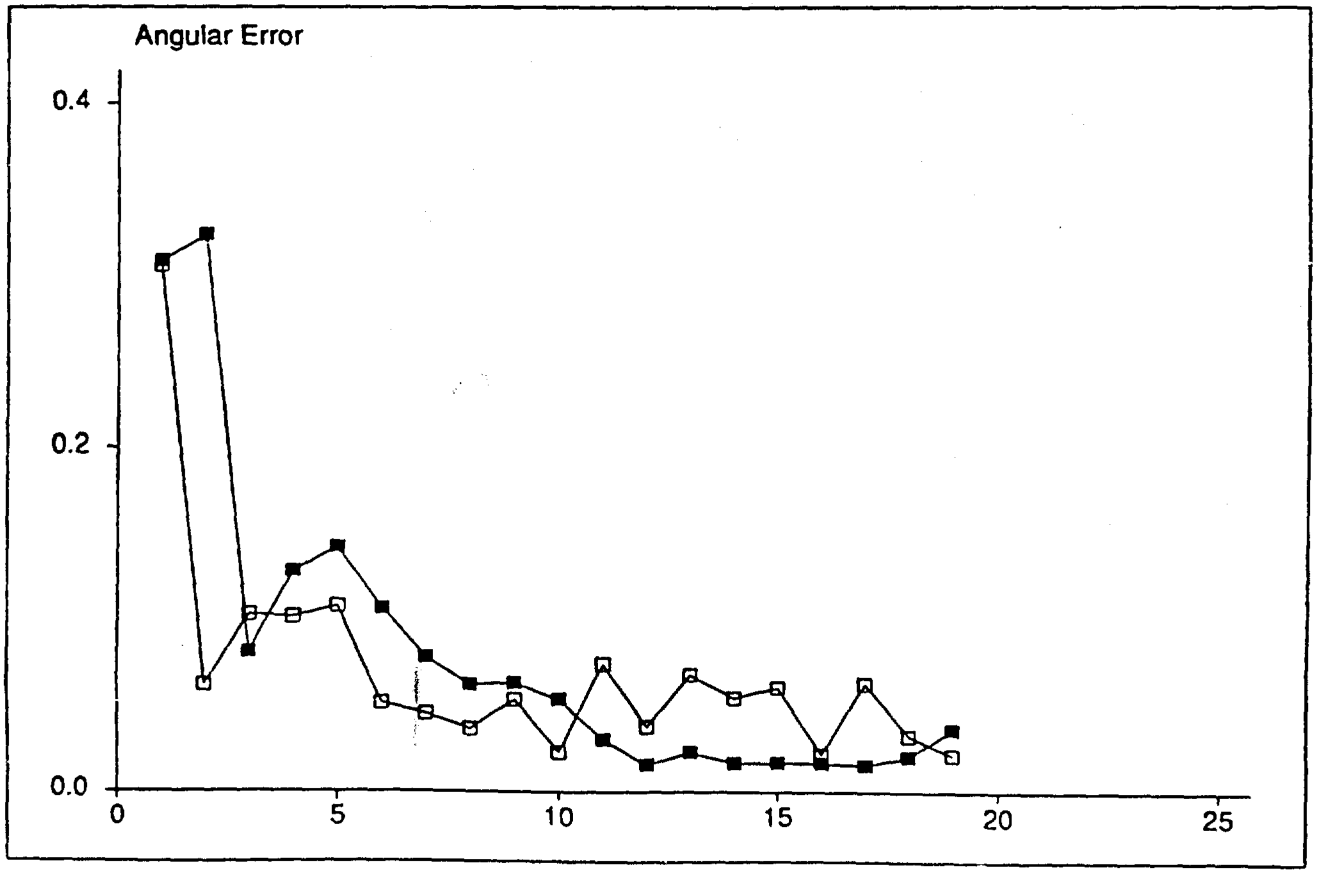

Figure 5a

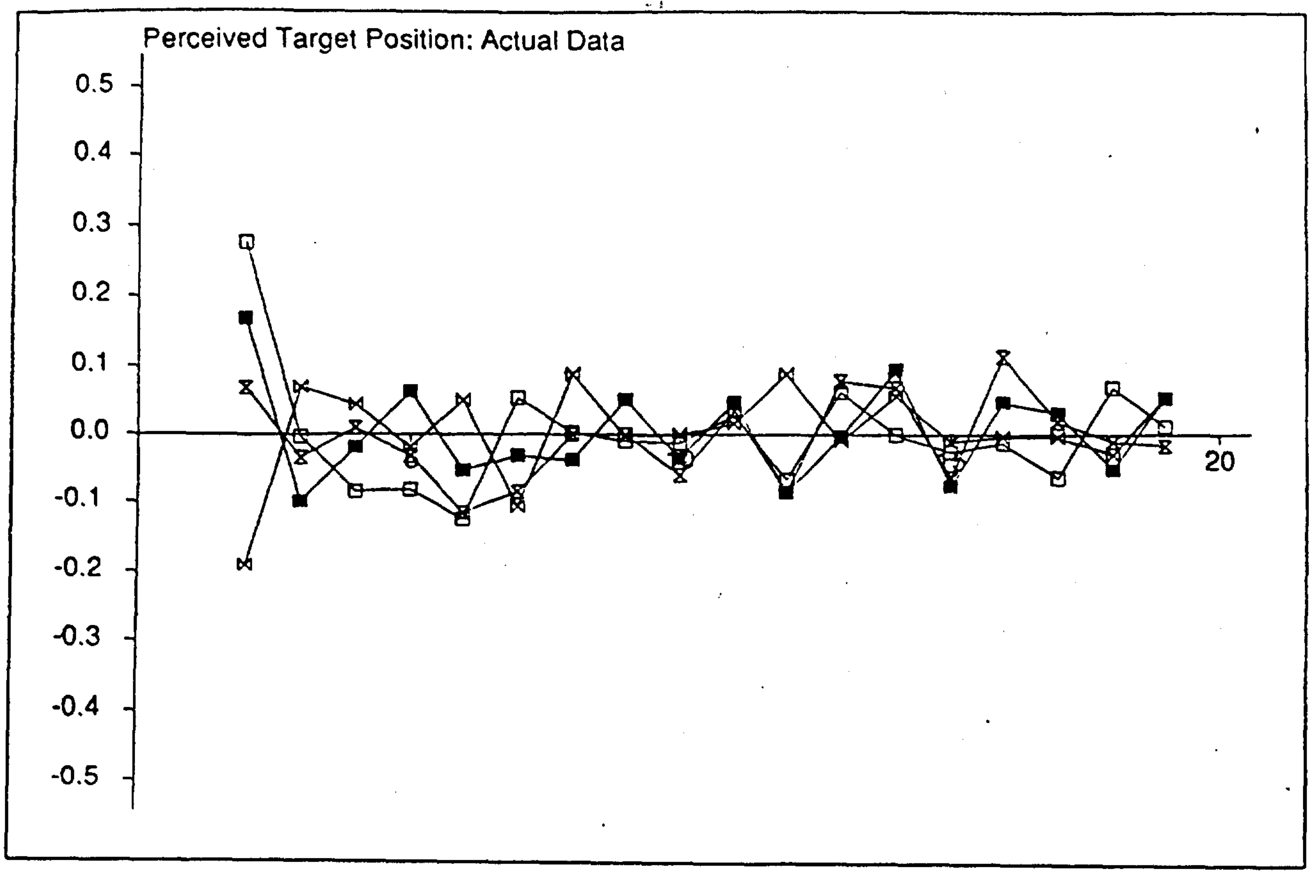

Figure 5b

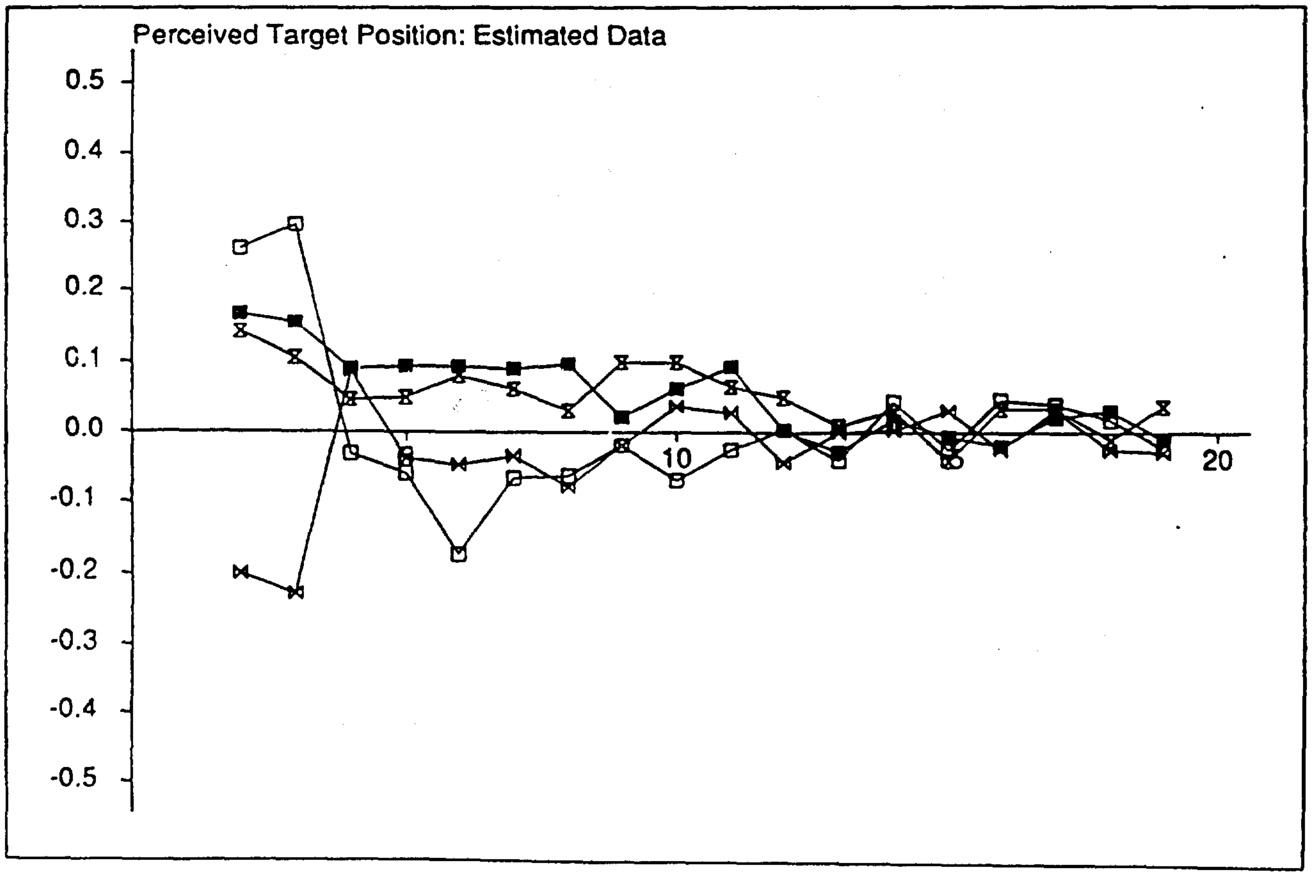

Figure 5c

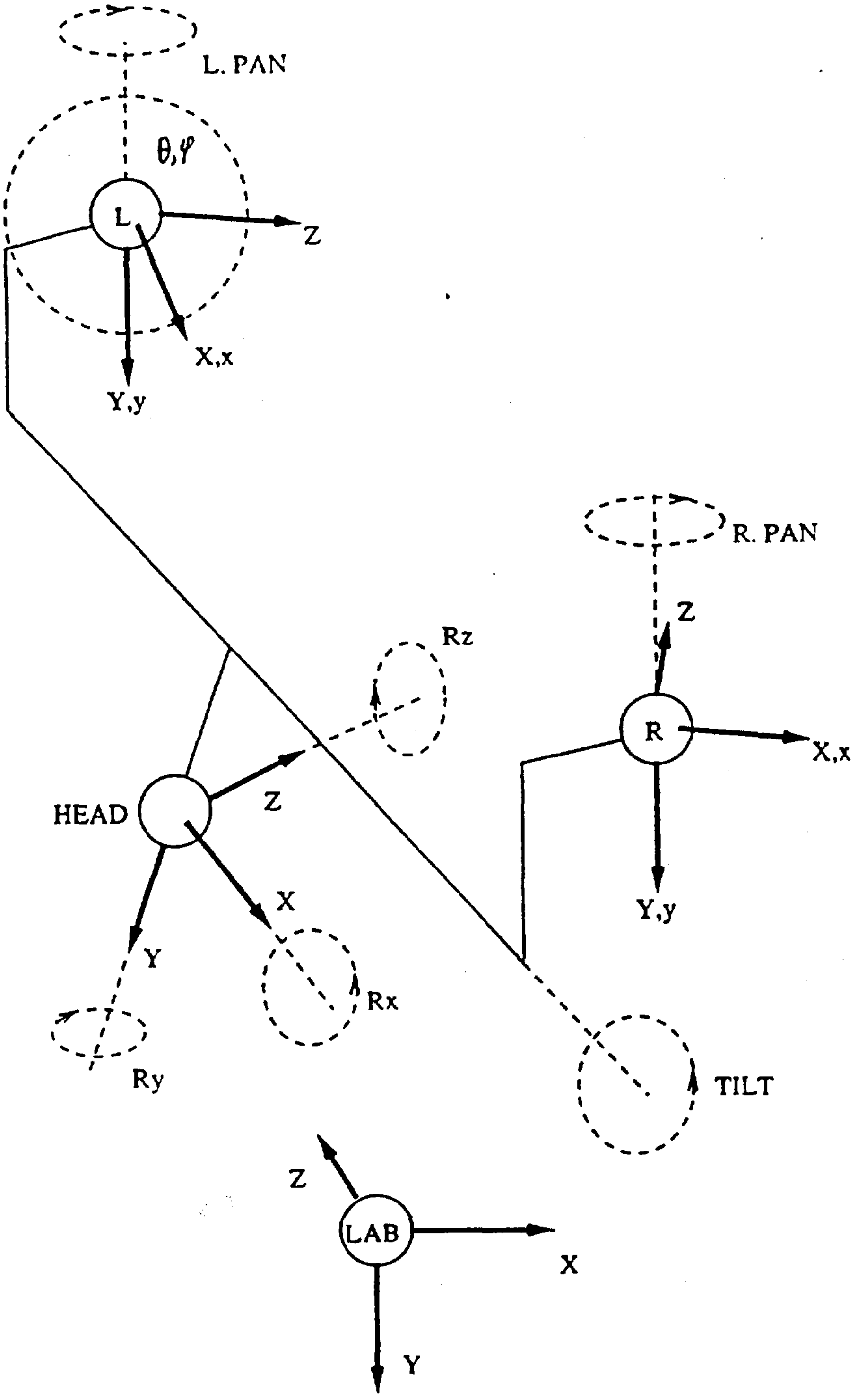

Figure 6

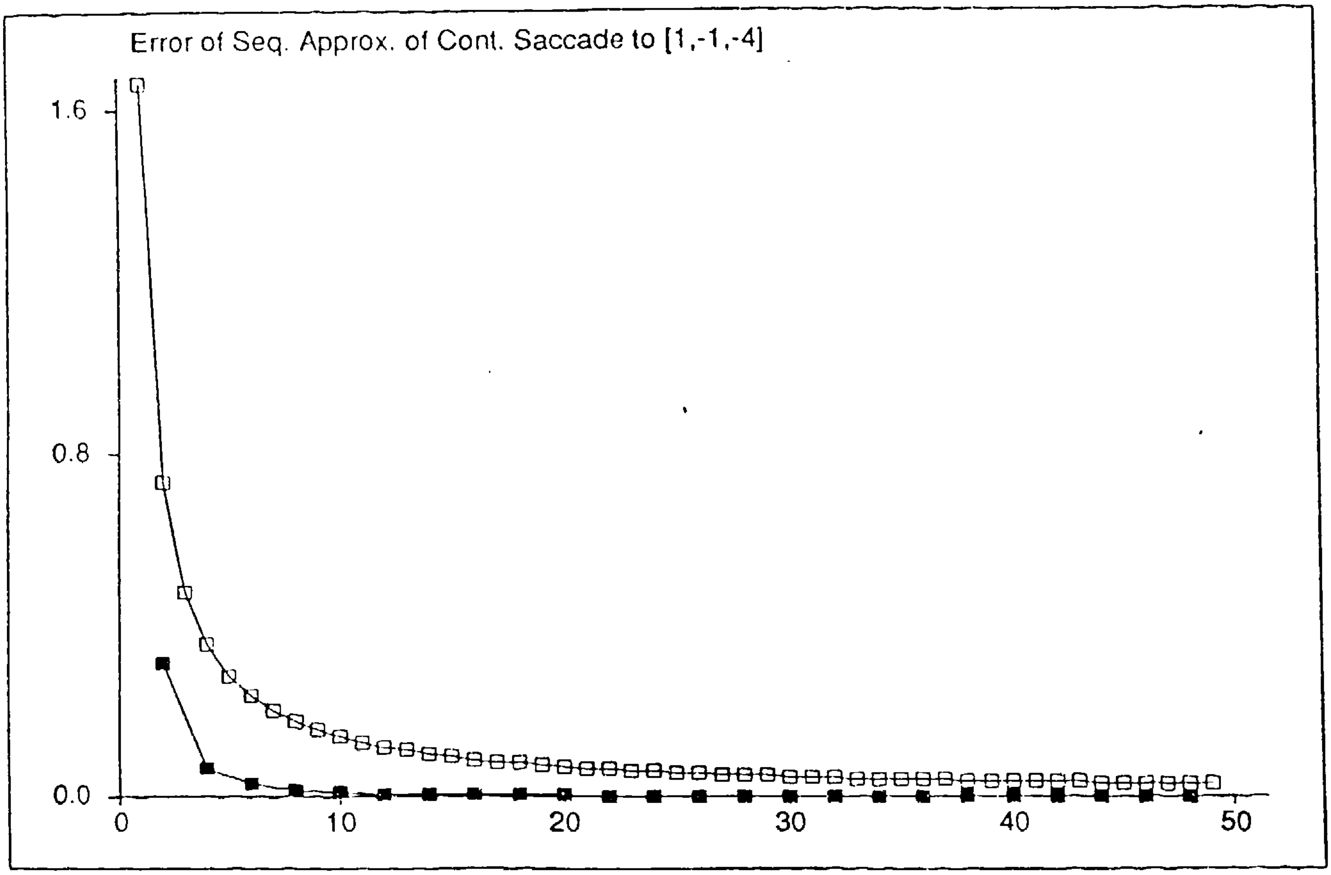

Figure 7

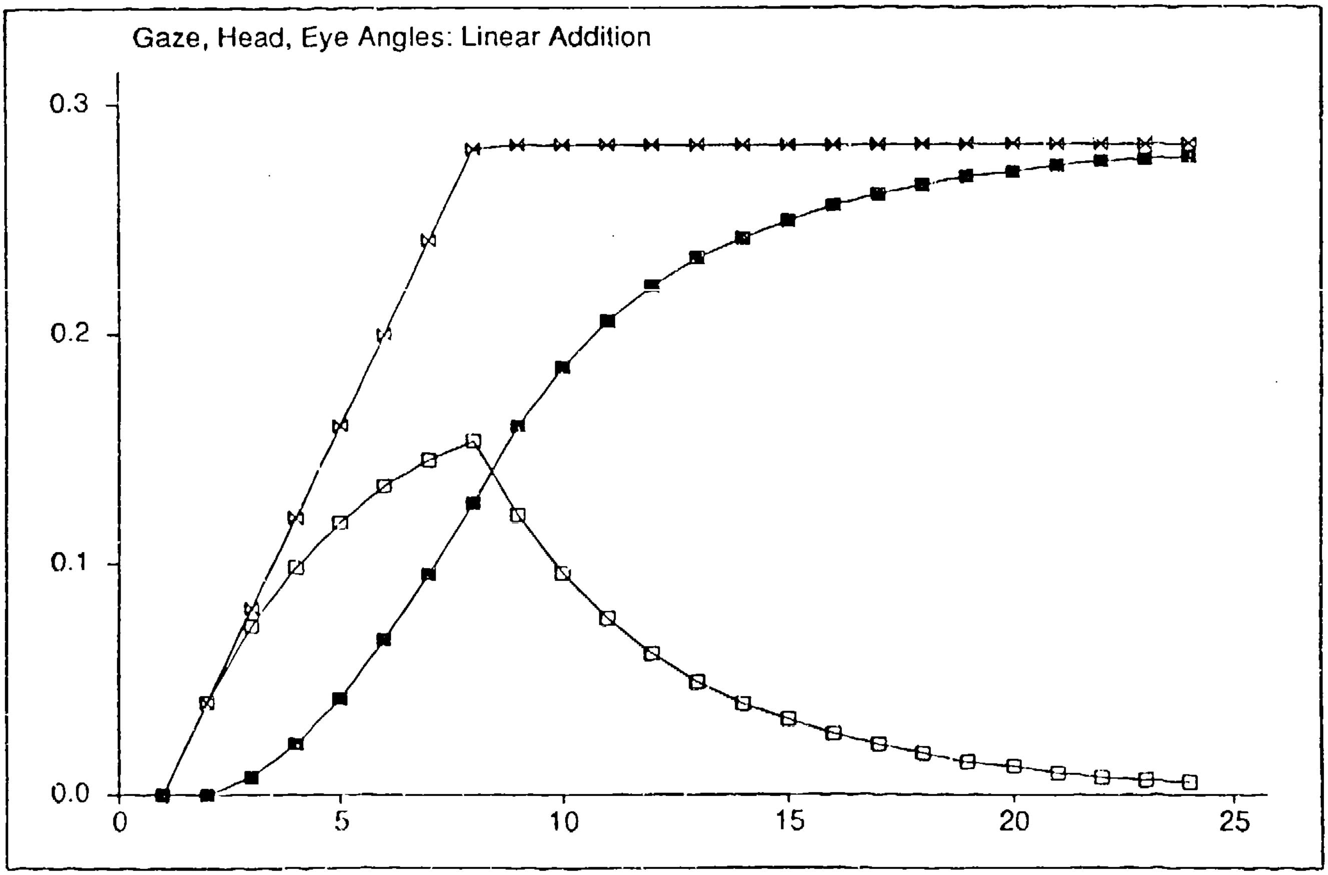

Figure 8

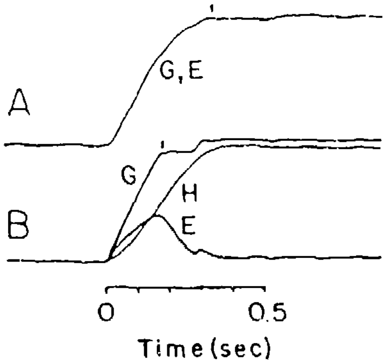

Figure 9a

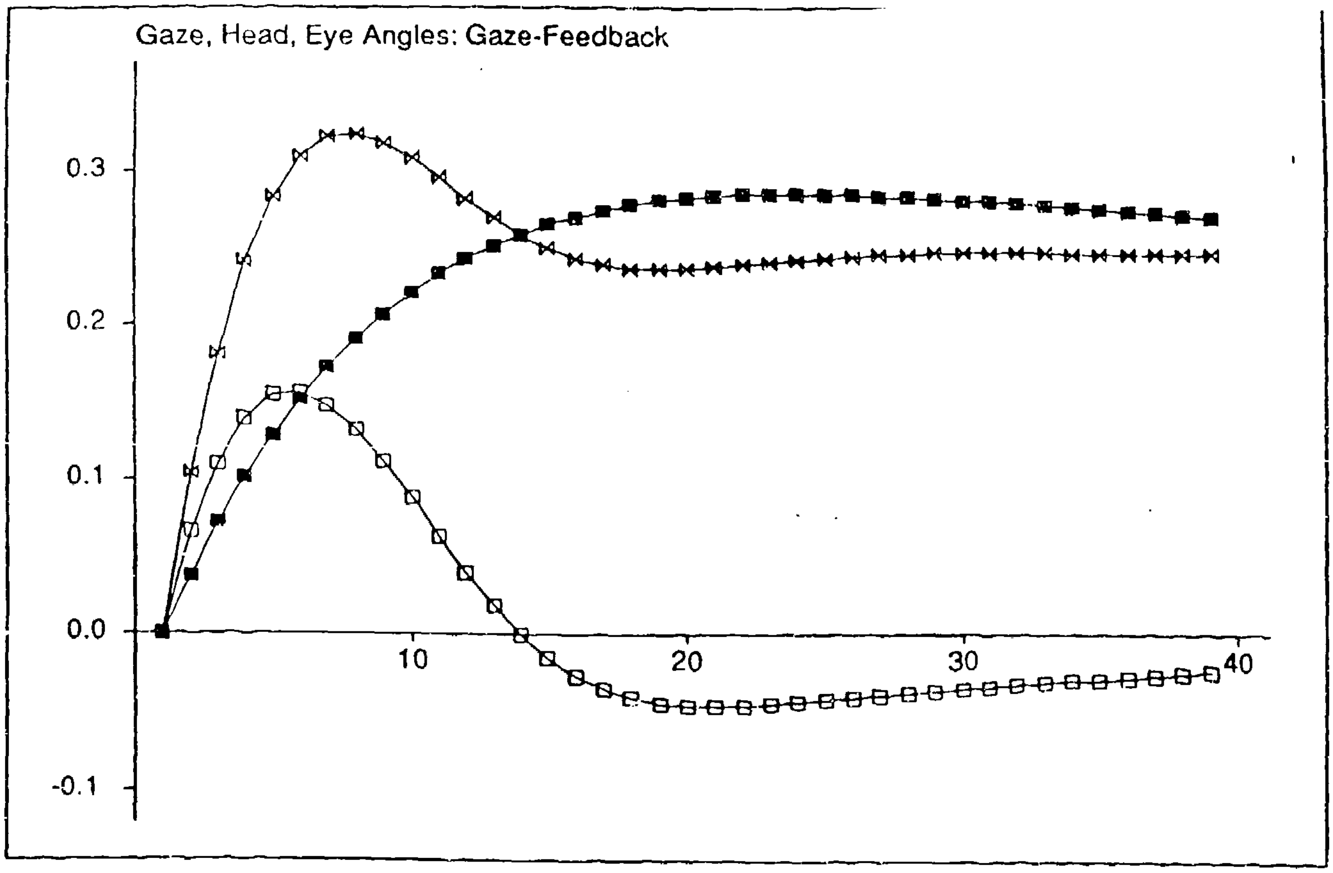

Figure 9b

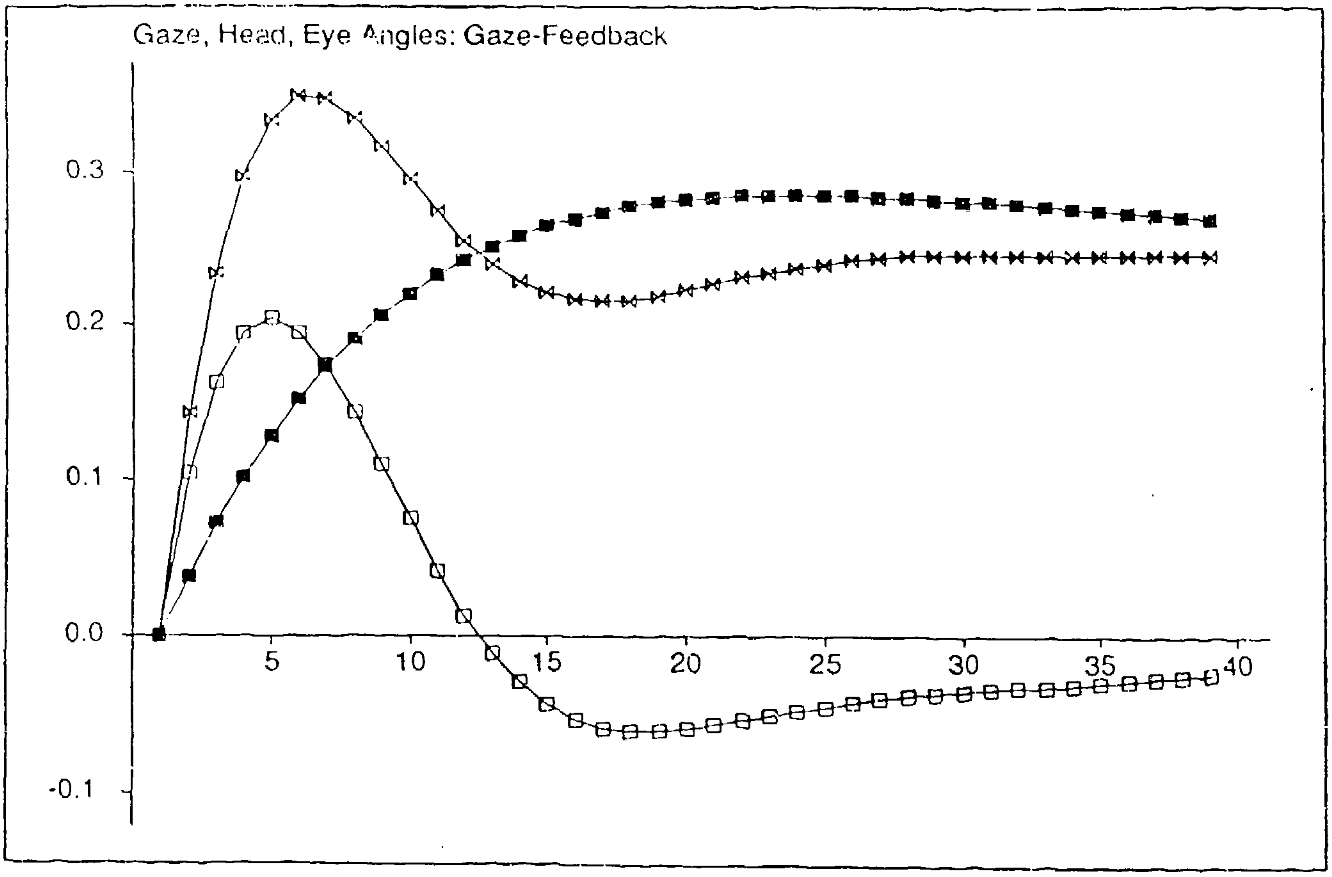

Figure 9c

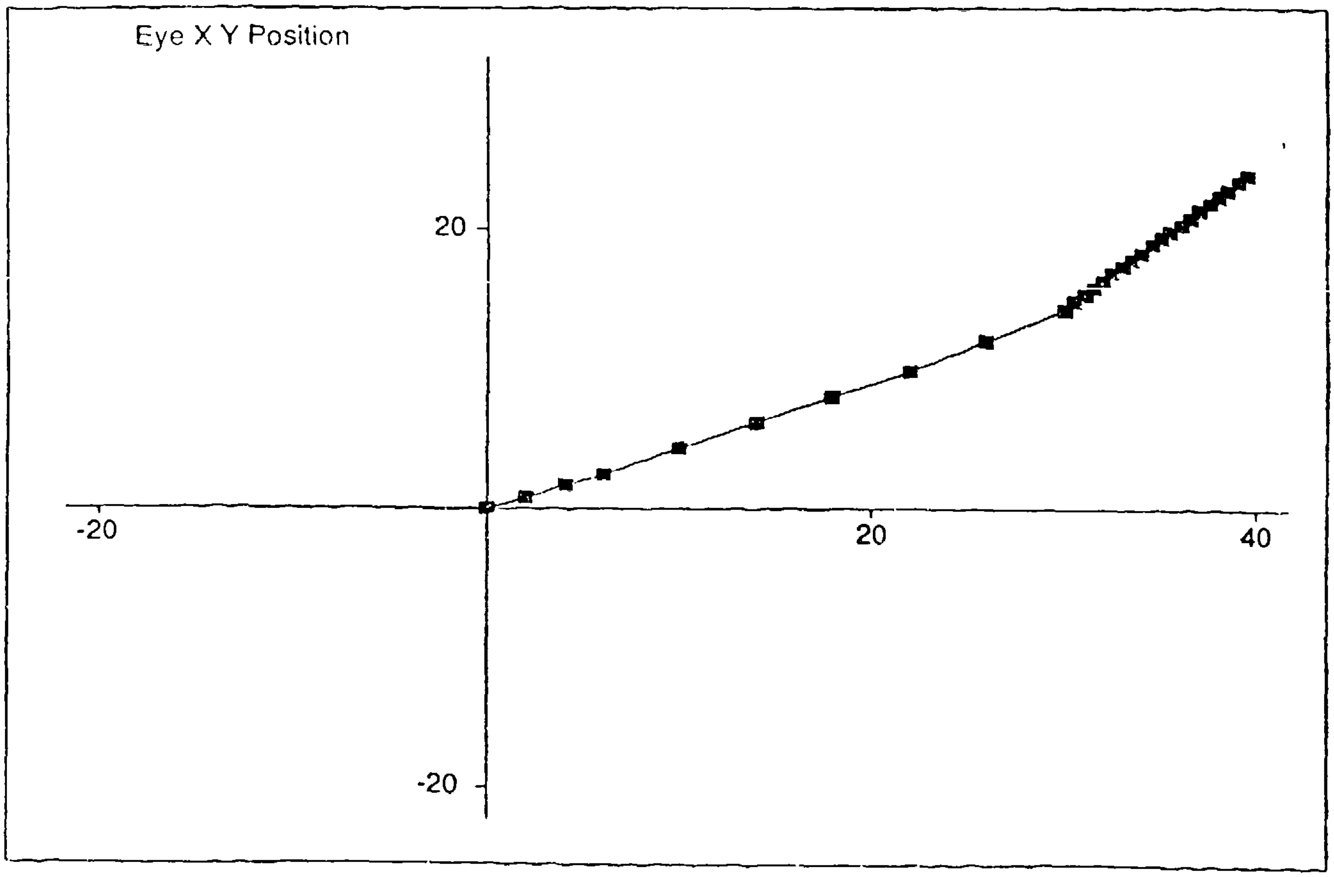

Figure 10a

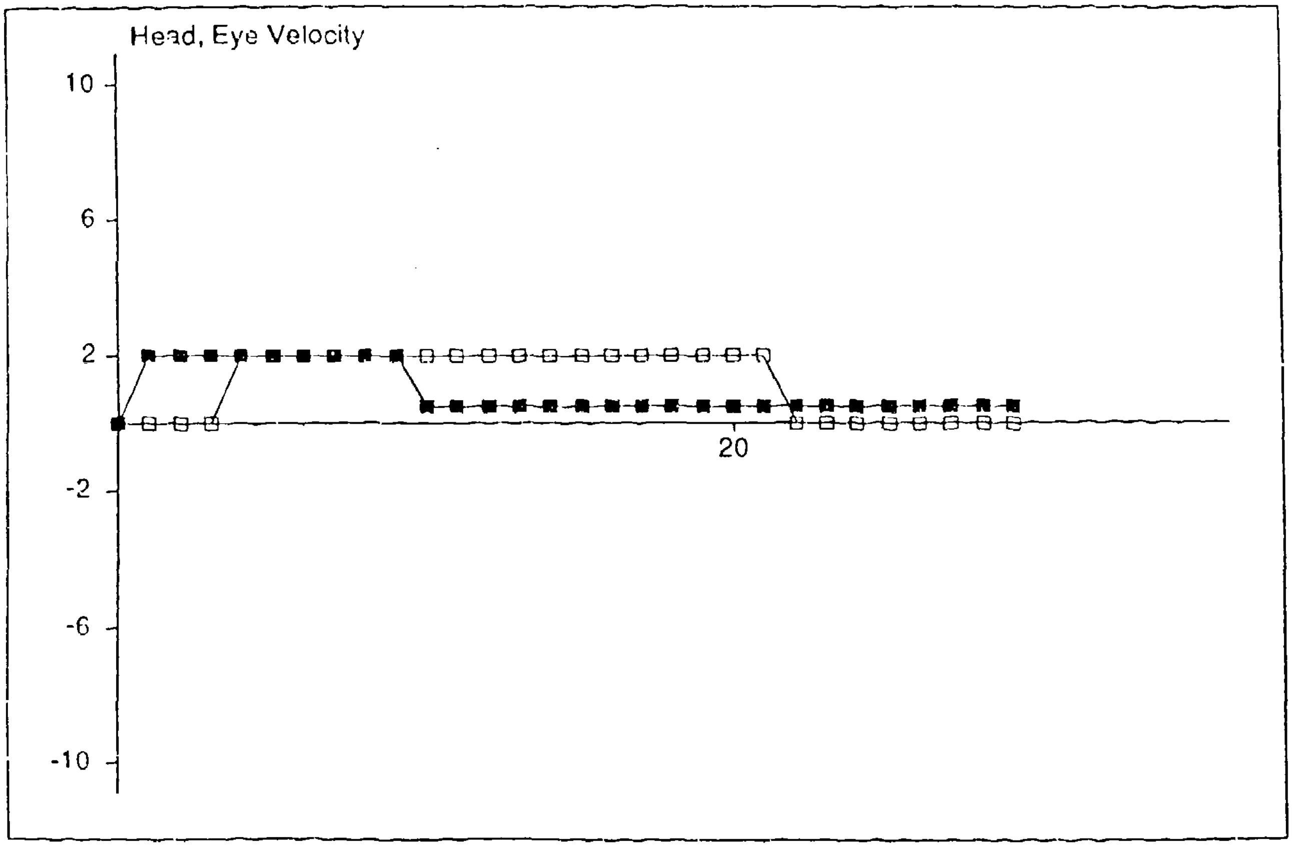

Figure 10b

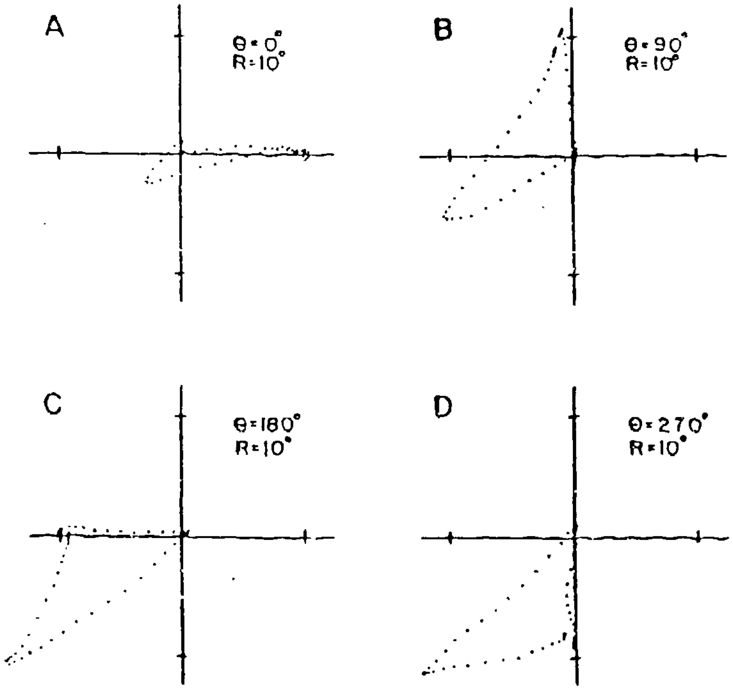

Figure 10c

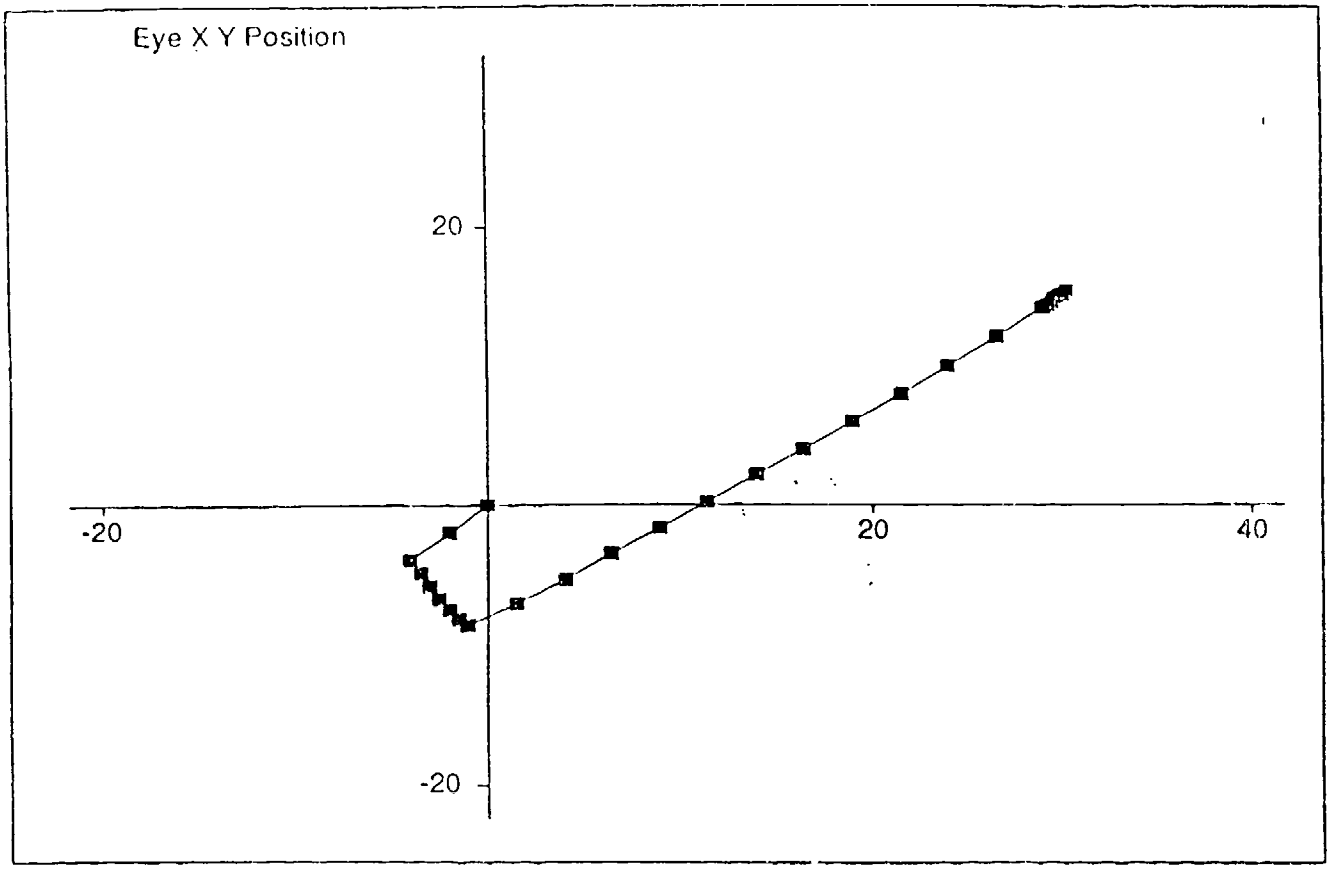

Figure 10d

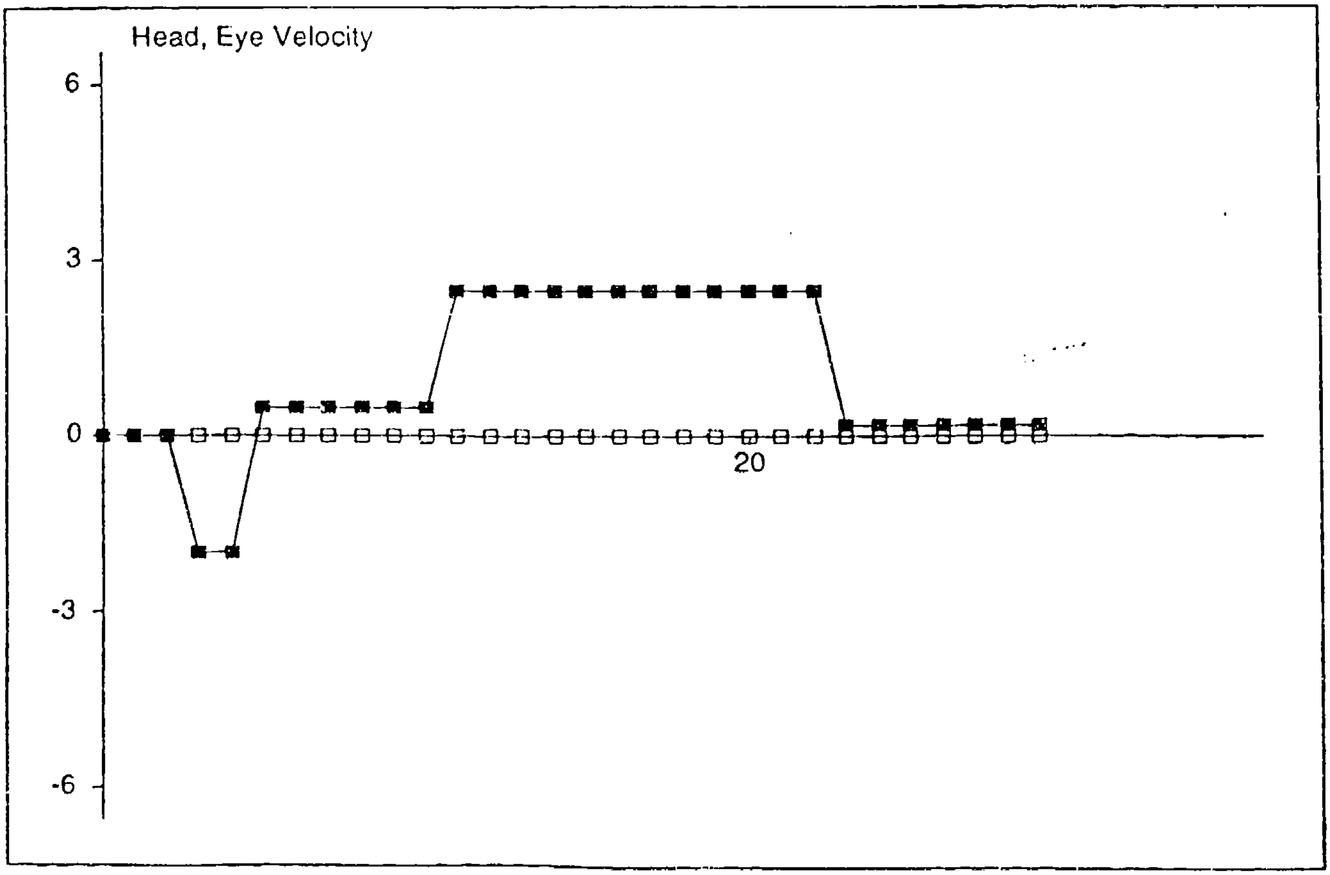

Figure 10e

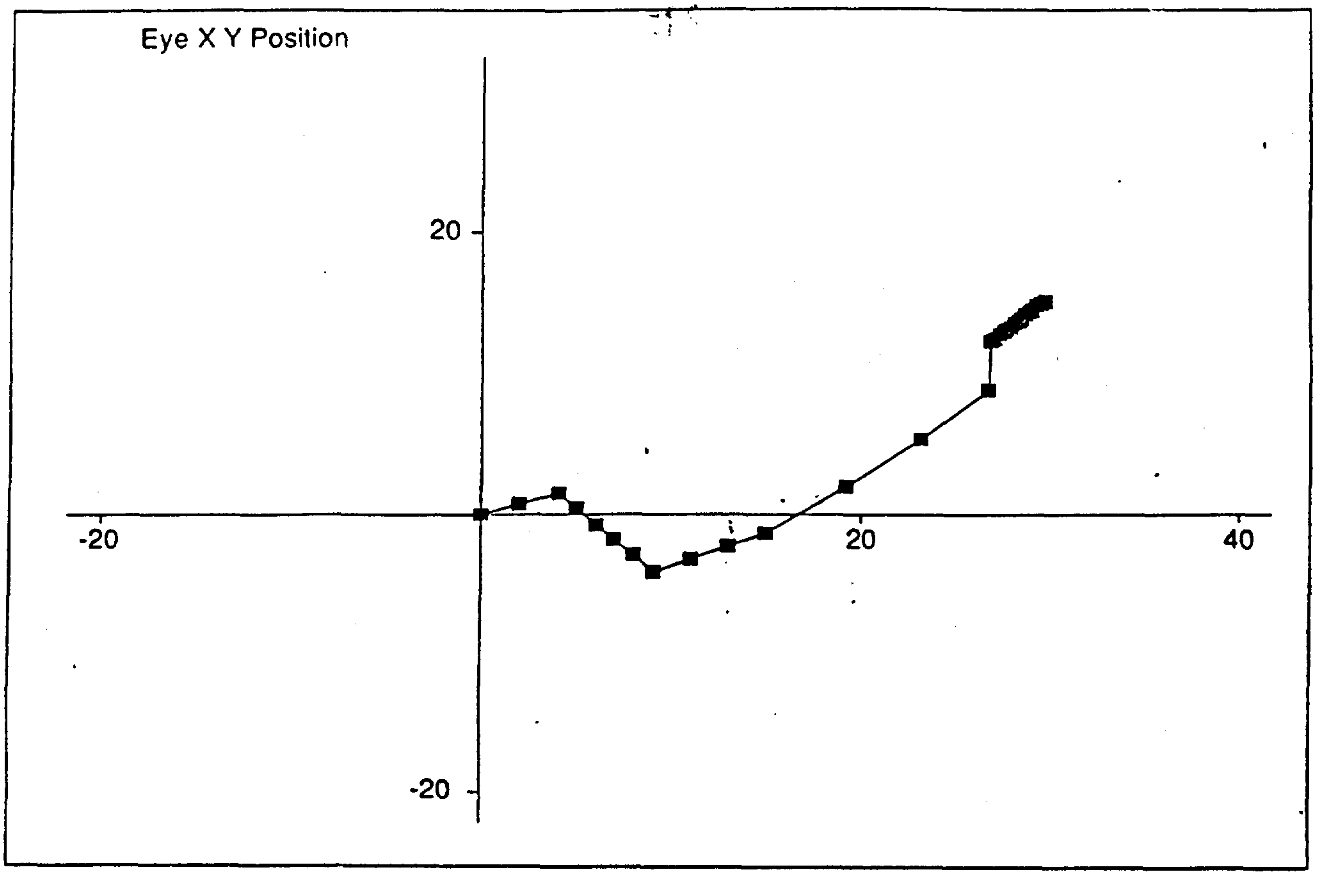

Figure 10f

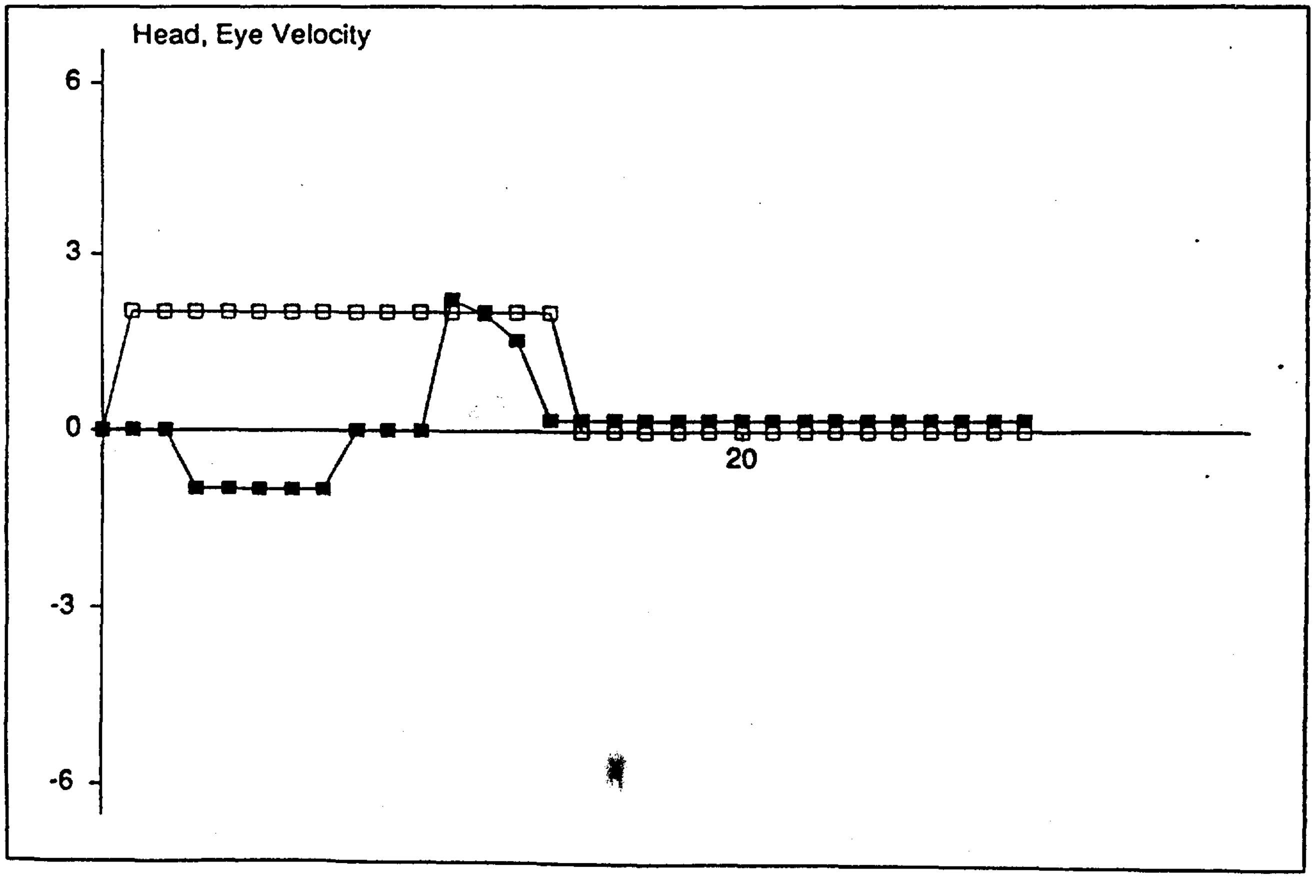

Figure 10g